

RADIO ENGINEERING HANDBOOK

KEITH HENNEY, EDITOR-IN-CHIEF

*Fellow, The Institute of Radio Engineers;
Author, "Principles of Radio," "Electron
Tubes in Industry"; Consulting Editor,
"Electronics"*

FOURTH EDITION

McGRAW-HILL BOOK COMPANY, INC.

NEW YORK TORONTO LONDON

1950

RADIO ENGINEERING HANDBOOK

**Copyright, 1933, 1935, 1941, 1950, by the McGraw-Hill Book Company, Inc.
Printed in the United States of America. All rights reserved. This book, or parts
thereof, may not be reproduced in any form without permission of the publishers.**

THE MAPLE PRESS COMPANY, YORK, PA.

PREFACE

In preparing new material and revising existing material for this Fourth Edition, the same guiding principles have been followed as in the First Edition in 1933. Throughout, the endeavor has been to create a comprehensive working manual of the radio science and to compile in a single volume concise information on each of the branches of radio engineering. In the years that have elapsed since the Third Edition, the war and all its technology have intervened; and much that existed in 1941 has been radically changed or has disappeared from the engineer's tool kits; and of course much that was only thought of—if at all—in that year is now standard practice. For example, the Third Edition does not mention wave guides. This volume has an entire chapter devoted to the subject.

Several chapters are totally new, for example, those on inductance and magnetic materials, wave propagation, electron tubes (much of which has never been published before), antennas, receiving systems, and radio aids to navigation. All other chapters have been thoroughly revised.

The first chapter in this edition, also new, represents a concise summation of the principles upon which the electrical communication art, radio particularly, is founded.

Much material from previous editions has been eliminated (the mathematical and mechanical tables, for example) not only to provide space for recent techniques but also to present a book economical in volume, weight, and cost. Nevertheless, this Fourth Edition is about 80 per cent larger than the Third Edition.

While there is much of what may be called fundamental background in this book, the emphasis has been on working practice rather than theory as a general concept of the purpose of a "handbook."

The engineer will find in this book many man-hours of effort compiled in the form of tables and curves or converted into concise English by the engineers, physicists, and teachers who have aided the editor in preparing this new edition.

KEITH HENNEY

NEW YORK, N.Y.
February, 1950

CONTENTS

PREFACE	v
-------------------	---

CHAPTER 1

BASIS OF RADIO COMMUNICATION, <i>by</i> BEVERLY DUDLEY.	1
---	---

Principles of electrical communication—Frequency range—Band widths—Psychological responses—Sound—Vision—Analytical approach to electrical communication problems—Fundamentals of electromagnetism—Electrostatics—Magnetostatics—Electrodynamics—The electromagnetic field—Electric circuits—Circuit elements—Circuit laws—Magnetic circuits—Electromagnetic waves—Units and dimensions.

CHAPTER 2

RESISTANCE, <i>by</i> JESSE MARSTEN.	83
--	----

General concepts—Units—Properties of materials—Skin effect—Resistor types—Voltage, high-frequency, temperature, noise characteristics—Resistor rating—Varistors—Thermistors—Standards and specifications.

CHAPTER 3

INDUCTANCE, <i>by</i> F. G. WEBER; MAGNETIC MATERIALS, <i>by</i> RAYMOND L. SANFORD	118
---	-----

Resistance effects—Distributed capacitance effects— Q —Inductance measurement—Inductor types—Calculation of inductance—Insulation, finish, shielding—Powdered-iron cores—Magnetic quantities and units—Magnetic materials—Magnetic testing—A-c and d-c measurements.

CHAPTER 4

CAPACITANCE, <i>by</i> E. L. HALL	163
---	-----

Dielectric materials—Capacitance calculation—Color codes—Electrolytic capacitors—Capacitor design—Standard capacitors—Capacitance measurements.

CHAPTER 5

COMBINED CIRCUITS OF L , C AND R , <i>by</i> W. F. LANTERMAN.	188
---	-----

Transients in LCR circuits—Pulsed circuits—Differentiating and integrating circuits—Steady-state currents—Resonance—Series and shunt impedance losses— LC tables—Oscillator tracking circuits—Coupled circuits—Band width and gain—Special LCR circuit applications—Transmission lines and networks—Delay networks—Pad design—Wave filters.

CHAPTER 6

ELECTRICAL MEASUREMENTS, <i>by</i> R. F. FIELD AND JOHN H. MILLER	254
---	-----

Units—Current measuring instruments—High-frequency current meters—Thermocouples—Rectifier meters—Power-level instruments—Voltage-measuring instruments—Resistance measurements—Moving-diaphragm meters—Elec-

tron-tube meters—Vacuum-tube voltmeters—Electron-stream meters—Wave analysis—Signal generators—Comparison measurements—A-c bridge measurements—T networks.

CHAPTER 7

POWER-SUPPLY SYSTEMS, *by* RICHARD C. HITCHCOCK 308

Characteristics and measurement of d.c.—Vacuum-tube rectifiers—Voltage doublers—Gas-tube circuits—Battery chargers—Dry metallic rectifiers—Transformer design—Filters—Filter-choke design.

CHAPTER 8

AUDIO-FREQUENCY AMPLIFIERS, *by* GLENN KOEHLER 346

Classification—Requirements—Design—Class A amplifiers—Coupling systems—Calculation of transformer constants—Push pull—Inverters—Direct coupling—Power amplifiers—Feedback—Gain control—Equalization—Amplifier measurements—Measurements of transformer impedance and iron-core reactance.

CHAPTER 9

RADIO-FREQUENCY AMPLIFIERS, *by* R. S. GLASGOW 390

Resistance-coupled r-f amplifiers—Tuned impedance-coupled amplifiers—Cascade amplifiers—Tuning stability—Grounded-grid amplifiers—Cathode followers—Neutralization—R-f power amplifiers—Modulated amplifiers—High-efficiency power amplifiers—Frequency multipliers.

CHAPTER 10

VACUUM-TUBE OSCILLATORS, *by* ROBERT I. SARBACHER AND DANIEL C. FELDER 416

Classification of oscillators—Feedback oscillators—Crystal-controlled oscillators—Magnetostriction—Tuned-filter circuits—Resistance stabilization—Bridge stabilization—Amplitude control—R-c oscillators—Dynatron—Beat-frequency oscillators—Relaxation circuits—Impulse generators—H-f oscillators—Magnetron and Klystron oscillators—Power-oscillator design.

CHAPTER 11

MODULATION AND DETECTION, *by* LESLIE F. CURTIS 461

Amplitude modulation—Pulse, phase and frequency modulation—Interference—Noise—Nonlinear circuits—Amplitude modulators—Pulse modulators—Multivibrators—Modulators for p.m. and f.m.—Converters and mixers—Detectors—Detectors of p-m and f-m waves—Limiters—Synchronized oscillators—Frequency multipliers.

CHAPTER 12

RADIO-WAVE PROPAGATION, *by* DALE POLLACK 523

Frequency-range classification—Characteristics of radio waves—Propagation by ground wave, sky wave, direct wave—Atmospheric effects—Ionosphere characteristics—Noise.

CHAPTER 13

ELECTRON TUBES, *by* G. D. O'NEILL 545

Electron emission—Cathode design—Filament design—Emission tests—Phototube cathodes—Secondary and field emission—Space-charge-limited currents—

Internal effects—Tube parameters—Multielement tubes—General tube design—Converters—Space-charge and transit-time effects—Beam modulated tubes—Klystrons and magnetrons—Traveling-wave tubes—Cathode-ray tubes.

CHAPTER 14

ANTENNAS, *by* EDMUND A. LAPORT. 602

Terminology—Current distribution in linear conductors—Low-frequency transmitting antennas—Scale models—Insulation problems—Directive antennas—Medium-frequency broadcast antennas—Arrays—H-f antennas—Rhombics—Long-wire antennas—V-h-f and U-h-f antennas—Pattern shaping—Lobe suppression—Antenna measurements.

CHAPTER 15

WAVE GUIDES AND CAVITY RESONATORS, *by* THEODORE MORENO. 684

Wave guides and transmission lines—Phase and group velocity—Circular wave guides—Elliptical guides—Characteristic impedance—Power-carrying capacity—Attenuation—Dielectric wave guides—Obstacles and discontinuities—Connectors—Launching waves in wave guides—Mode transformers—Tees and junctions—Standing-wave detectors—Attenuators—Impedance transformers—Cavity resonators—Echo boxes.

CHAPTER 16

LOUD-SPEAKERS AND ROOM ACOUSTICS, *by* HUGH S. KNOWLES 731

Speech, music, noise and hearing—Loud-speaker theory—Horns—Diaphragms—Motors—Efficiency-frequency characteristic—Baffles, enclosures, cabinets—Vents—Room acoustics—Reverberation—Power requirements—Objective loud speaker tests—Subjective tests.

CHAPTER 17

RECEIVING SYSTEMS, *by* CHARLES E. DEAN 775

Receiver classification—Methods of rating and testing—Noise factors—Performance of typical receivers—Noise in receivers—Loop-antenna receivers—R-f amplifiers—Cross modulation—Superheterodynes—Spurious responses—Converters—Push-button tuning—I-f amplifiers—Volume control—Automatic volume control—Tuning indicators—Squelch and noise-limiting circuits—Tone control—Automatic frequency control—Remote control and telemetering—Multiband receivers—Radio-phonograph combinations—Volume expansion and contraction—A-c-d-c receivers—F-m receivers—High fidelity—V-h-f receivers—Super-regenerators—Pulse receivers—Mobile, railroad, marine, amateur, point-to-point receivers—Military receivers—Radar, i-f-f and beacon systems—Direction finders—Automatic alarm receivers—Aircraft receivers—Panoramic, diversity, wide-band receivers—Antennas for broadcast reception—Wiring codes.

CHAPTER 18

RADIO BROADCASTING, *by* CARL G. DIETSCH 895

System elements—Audio range required—Volume range—Reference levels—Audio facilities—Microphones—Studio technique—Microphone placement—Volume indicators—Volume control—Program recording facilities—Wire, film and disk recording—Wire line characteristics—Radio facilities—Limiting ampli-

fiers—Monitoring equipment—Amplitude- and frequency-modulation transmitters—International broadcasting—R-f amplifiers—Modulation equipment—Carrier stabilization—Transmitter power supply—R-f harmonic suppression—Transmission line engineering—Antenna terminations—Transmitter adjustments—Field intensity measurements—Calculation of station coverage—V-h-f coverage.

CHAPTER 19

TELEVISION, *by* DONALD G. FINK. 994
Scanning and image analysis—Interlacing—Resolution—Video-signal wave form—Television cameras—Synchronization and shading-correction generators—Video amplifiers—Video modulation—Detection—Sync-signal separation—Picture tubes and associated circuits—Scanning and synchronization—Contrast and gradation—Color systems.

CHAPTER 20

FACSIMILE, *by* R. E. MATHES 1050
Scanning—Modulation—Photographic recorders—Synchronization—Propagation—Tape-facsimile systems—Operating standards—Facsimile applications.

CHAPTER 21

RADIO AIDS TO AVIATION, *by* HARRY DIAMOND. 1067
Trends and standards in aviation radio facilities—U.S. Government facilities—Direction-measurement equipment—Beacons—ILS—GCA—Direction finding—Air-transport company facilities—Navigation systems—Loran, Gee, Decca—Omnirange—Consol—DME—Shoran—Micro H—Accuracy evaluation—Altimeters—Present problems—Interrogation and transponder systems—Traffic-control systems—Noise problems—Precipitation static.

CHAPTER 22

CODE TRANSMISSION AND RECEPTION, *by* JOHN B. MOORE AND DAVID S. RAU. 1159
Standard codes—Signal characteristics—Radiotelegraph services—High versus low frequencies—Transmitting systems and equipment—Receiving systems and equipment—Diversity reception—Control methods and problems—Transcribing—Recorders and printers.

INDEX 1185

CHAPTER 1

BASIS OF RADIO COMMUNICATION

BY BEVERLY DUDLEY¹

Radio communication may be defined as the interchange of signals, symbols, intelligence, or subjective impressions between two or more points, employing electromagnetic waves as the medium of transmission. Engineering may be defined as the art and science by which natural forces and materials are utilized in structures, machines, or mechanisms for man's benefit. Radio engineering—that branch of engineering which deals with radio communication—may, therefore, be defined as the art and science by which natural forces and materials are utilized for man's benefit in mechanisms intended to convey the interchange of intelligence between two or more points through the use of electromagnetic waves as the transmission medium.

Both in theory and in practical application, a very large part of radio engineering may be regarded as a broadened and generalized field of electrical engineering although, in addition, some phases of radio engineering include other branches of physics such as acoustics, optics, meteorology, crystallography, and similar fields. Radio engineering practice differs from electrical engineering practice primarily in two important respects. The radio engineer is nearly always interested in transferring the maximum amount of power from the transmitting to the receiving ends of the system, whereas the electrical engineer is usually more interested in the transmission of power between two points with a maximum of efficiency. Radio engineering practice also differs from electrical engineering practice in the much greater range of frequencies employed and in the utilization of certain electrical phenomena whose effects are negligible at the comparatively low frequencies at which electric power is generated, distributed, and utilized. Radio engineering requires a wide range of frequencies in order that electro-

Table 1. Classification of the Electromagnetic Spectrum Used in Radio Communication

Carrier frequency, cps	Free-space wave- length, meters	Frequency-band designation	Principal communication uses
0-20 20-10 ⁴ 10 ⁴ -3 × 10 ⁴ 3 × 10 ⁴ -3 × 10 ⁵	∞-1.5 × 10 ⁷ 1.5 × 10 ⁷ -3 × 10 ⁸ 3 × 10 ⁸ -10 ⁴ 10 ⁴ -10 ⁸	Subaudible Audio (a.f.) Very low (v.l.f.) Low (l.f.)	Telegraphy Reproduction of speech and music Long-wave fixed stations Point to point, transoceanic, govern- ment, direction finding
3 × 10 ⁵ -3 × 10 ⁶	10 ⁴ -10 ⁵	Medium (m.f.)	Broadcasting, ship-harbor, government, amateur, police
3 × 10 ⁶ -3 × 10 ⁷	10 ³ -10	High (h.f.)	Coastal ship, international broadcast- ing, government, aviation, amateur
3 × 10 ⁷ -3 × 10 ⁸	10-1	Very high (v.h.f.)	Police, government, relay broadcast, television, coastal and ship-harbor, amateur
3 × 10 ⁸ -3 × 10 ⁹	1-0.1	Ultra-high (u.h.f.)	Government, air navigation, facsimile, television, citizens' radio, meteorologi- cal, amateur
3 × 10 ⁹ -3 × 10 ¹⁰	0.1-0.01	Super-high (s.h.f.)	Government, navigation aids, fixed and mobile, amateur
3 × 10 ¹⁰ and more	0.01 and less	Experimental

¹ Massachusetts Institute of Technology, Cambridge, Mass.

magnetic waves may be produced practically and also in order that appreciable amounts of intelligence (considerable detail) may be imparted to the electrical system in the communication process. The range of frequencies that the radio engineer may encounter in his professional work is given in Table 1, in which the designations for the r-f portion of the spectrum are those established Mar. 2, 1943, by the Federal Communications Commission (FCC).

Modern radio communication rests on two fundamental bases: (1) the analysis, formulation, and manipulation of the intelligence to be conveyed, in a form suitable for transmission by radio means, and (2) the fundamentals of those branches of physics (primarily electromagnetism) by which electromagnetic waves may be produced, modulated in accordance with the intelligence to be transmitted, propagated through space, and received at a distant point where the intelligence is extracted. The proper combination of both of these topics provides the foundations for radio communication engineering.

PRINCIPLES OF ELECTRICAL COMMUNICATION

1. General Problem of Electrical Communication. Communication is carried on by the successive selection and arrangements of prearranged or understood physical symbols. In a manner that depends upon the properties of the communication system, such signals are produced by the sender and are brought to the attention of the receiver. The proper selection and sequential arrangement of such physical symbols or signals are capable of conveying intelligence in accordance with the meanings attached to the symbols. Since a sequence of symbols is required to convey information, it is evident that one of the characteristics of intelligence is that it may be expressed as a single-valued function of time. From a practical point of view this is a most important result for the communication engineer, for it is this characteristic property of intelligence that permits information to be transmitted economically by electrical means between distant points.

In telegraphy the different letters that spell out the words of the intelligence to be conveyed are resolved into dot and dash impulses. Each symbol for a letter, numeral, or punctuation mark differs from all others by the distinctive and unique combination of the active and inactive periods of each dot-dash arrangement. In telephony, intelligence is conveyed by uttering, syllable by syllable, the words of the sentences making up the intelligence. In speech only one value of sound pressure is produced at each instant of time, and thus speech may be regarded as a single-valued function of time. The same may be said for music because, although many players may combine their tones to produce a highly complex sound, at any instant the resultant sound pressure is a single-valued function of time.

When visual images are to be transmitted by electrical means, it becomes necessary to resolve the image (seen subjectively as a two-dimensional area by each eye) into a single-valued function of time. Usually the required space-time conversion is achieved by neglecting the fact that the two eyes see slightly different images as a result of their separation by the interpupillary distance and by employing only a single image (as seen by one eye) to convey the visual information. The time-space conversion of the image is achieved by dividing the visual image into comparatively minute elements, each of which may be considered to be of uniform visual characteristics, and then effectively converting the picture elements into a single-valued function of time by a suitable process of scanning in which the picture elements are progressively selected according to a prearranged and orderly sequence and in which some visual characteristic of each picture element is converted into a corresponding electrical quantity.

No matter what type of signal is to be transmitted between two points, the precision of information depends upon what other symbol sequences might have been chosen; hence, the number of sequences required to convey a desired amount of information may be considered as a quantitative measure of information, having the important advantage that it is objective and completely free from physiological factors. In order that the measure of information may be of practical engineering value, however,

it should be of such character that the amount of information conveyed is proportional to the number of selections. The number of possible sequences is, therefore, not suitable for use directly as a measure of information.

If the total amount of information conveyed is proportional to the number of sequential selections, the rate of conveying intelligence is proportional to the number of elements selected per unit of time or to the frequency of selection. In the simplest case a cycle of variation may correspond to the selection of a single symbol, but since each signal (e.g., the letters of the Morse code) may be more complex than that by which the simplest element of intelligence may be expressed, several cycles of variation may be required for the selection of each symbol. For a given rate of intelligence transmission the frequency of variation (or the frequency of transmission) is proportional to—and is usually several times greater than—the frequency of selection of the intelligence symbols.

2. Frequency Range Related to Amount of Intelligence Transmitted. To be useful, each symbol expressing a certain amount and kind of intelligence must be different from every other symbol expressing a different amount and kind of intelligence. Some signals or symbols will be more complicated or more complex than others and will require a greater frequency of transmission than simpler signals. It follows that the transmission of intelligence, i.e., the selection of symbols, at a reasonably uniform rate is characterized by a range or band of frequencies of transmission. The range of transmission frequencies normally required for several different kinds of communication systems is given in Table 2.

Table 2. Transmission Band Widths Required for Various Types of Radio-communication Services

Kind of Service	Transmission Band Width, Cps
Manual keying telegraphy.....	0-15
Machine keying telegraphy.....	0-100
Facsimile picture transmission (8- × 10-in. picture, 100 lines/in., transmitted in 10 min)	0-1,500
Speech telephony.....	250-2,800
Ordinary line for broadcast telephone transmission.....	50-5,000
High-quality line for broadcast telephone transmission.....	40-8,000
Highest quality reproduction of speech and music.....	20-20,000
Low-definition television (180 lines, 24 frames/sec).....	24-425,000
High-definition television (525 lines, 30 frames/sec).....	30-4,500,000

We are led to an important principle of communication of great practical value enunciated by Hartley:

The maximum rate at which information may be transmitted over a system whose transmission is limited to frequencies lying in a restricted range is proportional to this frequency range. From this it follows that the total amount of information which may be transmitted over such a system is proportional to the product of the frequency range which it transmits by the time during which it is available for the transmission.¹

3. Fundamental Steps in Electrical Communication Systems. No matter how simple or how complex the communication system may be, the general problem of electrical communication may be resolved into four distinct operational steps:

1. The *formulation* of intelligence—expressed by means of signals or symbols—in a manner capable of producing subjective sensations in the mind of another person. This first step is common to all communication systems whether electrical or not. In radio-communication systems the most representative types of signals are the dot-dash signals of the Morse code for telegraphy, voice signals for telephony, and the scanning signals for picture transmission.

2. The *conversion* of the intelligence to be transmitted into electrical energy of such characteristics as to be capable of conveying information over the desired circuit or through the selected transmission medium. In radio-communication systems the band of (modulation) frequencies to convey the intelligence is superimposed upon a higher (carrier) frequency in a process of modulation to produce a band or channel of frequencies capable of passing through the transmission medium and containing the appropriate information.

¹ Hartley, R. V. L., *Transmission of Information, Bell System Tech. J.*, 7 (3), 535, July, 1928.

Each such modulated signal is identified and separated from all others on a carrier frequency basis, usually by assigning a separate carrier frequency band or channel for each station or service.

3. The *transmission* of the electrical form or representation of the intelligence from one point to another by electrical means and its reception at a distant point. In radio-communication systems the electromagnetic waves produced at the transmitting station constitute the transmission medium. The desired modulated carrier wave may be selected to the exclusion of other waves by resonating the receiving system to the appropriate carrier frequency channel or by other means.

4. The *retranslation* of the received electrical representation of the intelligence or information, by a process of demodulation or detection, into a form of energy that is capable of affecting one or more of the senses of the recipient in such a way as to produce a subjective sensation essentially the same as that of which the original intelligence formulation was capable. The demodulation process is one in which the original electrical representation of the intelligence is extracted from the modulated carrier. The electrical form of the signal may then be converted into aural or visual images simulating those produced at the transmitting end of the system.

The foregoing considerations permit the establishment of the fundamental concepts in quantitative form for any kind of electrical communication system. If i represents the intelligence, e the electrical form of energy corresponding to intelligence, and t time, and if the subscript 1 designates those functions occurring at the transmitting end of the system whereas the subscript 2 indicates those taking place at the receiving end of the communication system, then the steps in any electrical communication system may be expressed as

$$i_1(t) \rightarrow e_1(t) \rightarrow e_2(t) \rightarrow i_2(t)$$

in which the arrows indicate the direction in which the process proceeds. This notation merely expresses, in symbolic form, the four steps enumerated above and illustrated graphically in Fig. 1.

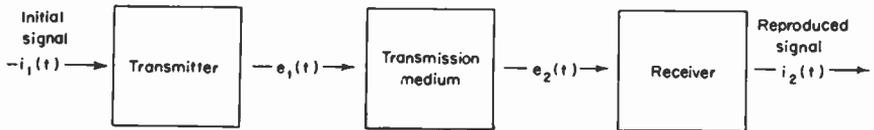


FIG. 1. Block diagram illustrating the fundamental steps involved in any system of electrical communication.

4. Perfect Transmission. The transmission and conversion of energy in a communication system must be carried out in such a way that, as a whole, the system is free from distortion or defects that prevent it from re-creating essentially the subjective sensations or impressions originally produced at the transmitting end. Distortion or other defects are most easily expressed in terms of deviations from perfect or ideal transmission.

For perfect (ideal) transmission the intelligence function re-created at the receiving end must be identical in all respects to that originally produced at the transmitting end. The condition imposed on an ideal communication system may thus be expressed quantitatively by the relation

$$i_2(t) = i_1(t)$$

The extent by which $i_2(t)$ differs from $i_1(t)$ is a measure of the over-all deviation from ideal transmission.

The requirement for ideal transmission, as given here, is more stringent than is usually necessary or economically feasible, and ideal transmission is seldom, if ever, encountered in practice. The concept of ideal transmission is exceedingly useful in establishing a criterion by which the performance of any communication system may be judged.

For engineering purposes the concept of distortionless transmission is more useful than that of ideal transmission. For distortionless transmission the communication system must reproduce, at the receiving end of the system, the same wave form (*i.e.*, intelligence-time function), as that originally produced at the transmitting end. Distortionless transmission allows for the finite time of transmission between the transmitting and the receiving points; usually it also allows for a change of scale or magnitude. Distortionless transmission differs from ideal transmission in that the sensation-perceived energy at the receiving end may be greater than, equal to, or less than the corresponding energy at the transmitting end. The received signal is also subject to a time displacement or time delay since electrical transmission does not occur instantaneously. This delay is not taken into account in ideal transmission as here defined.

5. Need for High Frequencies. In order that intelligence-bearing signals may be propagated by means of electromagnetic waves, it is necessary that the electric and magnetic components of the electromagnetic field interact with one another to support and sustain electromagnetic waves. This condition requires the generation of displacement currents of appreciable magnitude, and these are produced only when the carrier frequency is comparatively high. Practical radio communication depends upon the use of frequencies greater than (and usually very much greater than) about 10,000 cps.

An electromagnetic wave varying at the rate of 10,000 cps or higher and used to convey intelligence through space is called a "carrier" wave. Of itself the carrier wave has no ability to convey intelligence. To convey intelligence, some property or characteristic of the carrier wave must be varied or modulated in accordance with the information to be transmitted. The proper combination of the carrier and the signal waves, such that the carrier is able to convey intelligence through the transmission system, is called a "modulated" carrier wave.

Although the unmodulated carrier wave may be generated at a constant, single frequency, the modulated carrier is broadened into a band of frequencies called a "carrier channel." To overcome technical limitations, the frequency band width of the modulated carrier should be a small fraction of the unmodulated carrier (mean) frequency. This limitation requires very high carrier frequencies where, as in television, a considerable amount of detailed information must be transmitted in a short interval of time, or where the properties of the system necessitate reasonably large excursions from the mean carrier frequency, as in frequency modulation.

The ratio of the maximum transmission frequency to the (mean) carrier frequency for some typical classes of radio service is given in Table 3. These data show that the transmission band width is, at most, but a few per cent of the mean carrier frequency.

Table 3. Comparison between Transmission Band Widths and Carrier Frequency for Different Classes of Radio Service

Type of service	Typical carrier frequency, kc	Typical transmission band width, cps	Ratio of maximum transmission frequency to carrier frequency
Long-wave telegraphy.....	100	0-100	0.001
Aviation telephony.....	200	100-3,000	0.015
Broadcasting.....	1,000	50-5,000	0.005
Police radio (voice) telephony.....	2,000	100-3,000	0.0015
Tone-modulated telegraphy.....	3,500	1,000	0.00026
Facsimile picture transmission.....	45,000	0-3,000	0.00007
Frequency modulation.....	105,000 ± 75	30-15,000	0.00014
Television.....	180,000	30-4,500,000	0.025

6. Functional Elements of Typical Radio System. The general principles of an electrical communication system may be applied to the particular case of ampli-

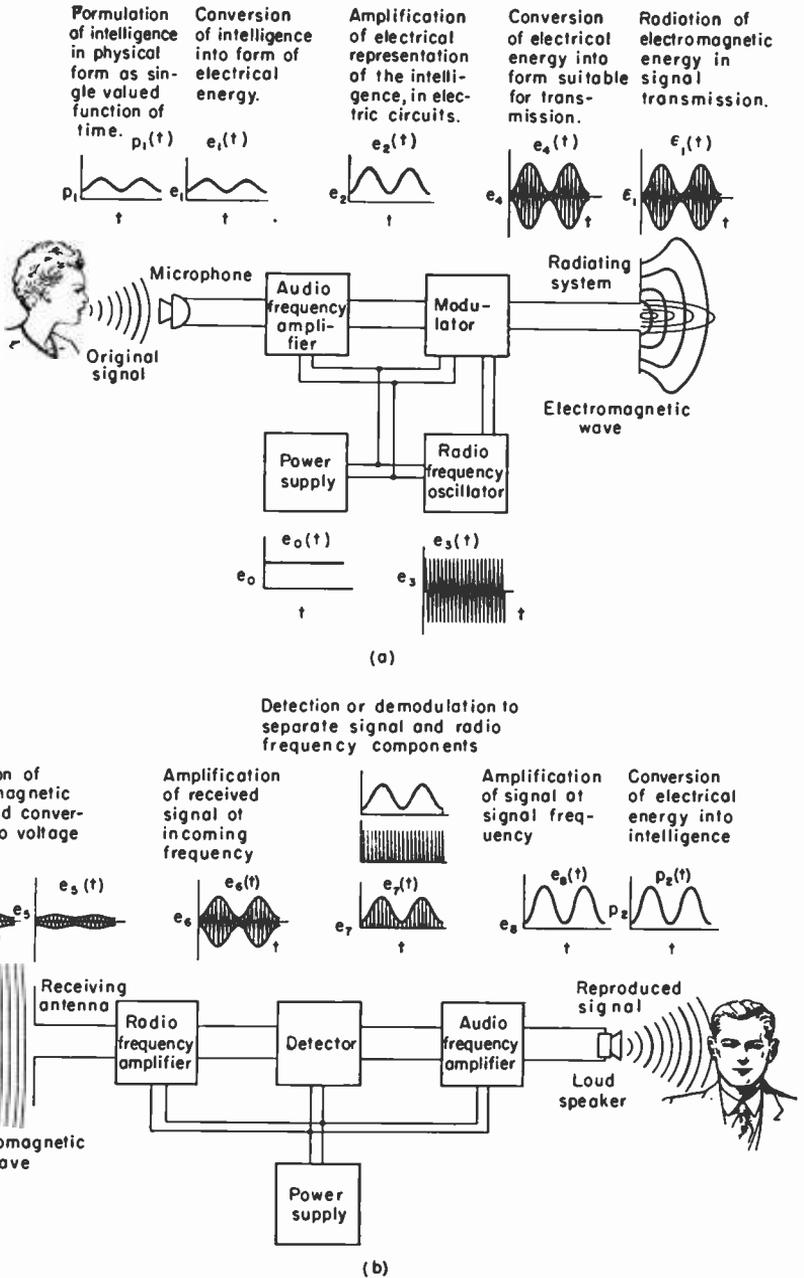


FIG. 2. Block diagram illustrating the general principles of an electrical communication system applied to the particular case of amplitude-modulated radiotelephony.

tude-modulated radiotelephony, for example, as indicated by the block diagram of Fig. 2.

In the transmitting end of the system, spoken words establish air-pressure variations, $p_1(t)$, which are converted by the microphone into variations of electrical energy, $e_1(t)$. The electrical form of the intelligence is then amplified, $e_2(t)$, and combined with a locally generated carrier oscillation, $e_3(t)$, whose output frequency corresponds to that of the assigned carrier channel. The result of this combination or modulation is the modulated carrier voltage (or current), $e_4(t)$, which, by means of an open-circuit radiating system, is converted into a wave-producing electromagnetic field, $\varepsilon_1(t)$.

From the transmitting radiating system, the modulated electromagnetic waves are radiated into space. In transmission, the field intensity of the wave is attenuated, and the wave reaches the receiving station with energy, $\varepsilon_2(t)$, greatly diminished from that which left the transmitting station. Although the radiating system at the transmitter may direct the major portion of the radiated energy into certain desired directions, it is a characteristic of radio communication systems that there is no control over the behavior of waves during transmission. In this respect radio communication differs from those systems in which the energy between transmitting and receiving ends is guided by conductors.

The electromagnetic wave at the receiving end, $\varepsilon_2(t)$, is converted at the receiving antenna into a voltage, $e_5(t)$, which is usually of very small magnitude. In tuned-radio-frequency (t-r-f) receivers the incoming voltage is amplified at the carrier frequency, $e_6(t)$. The carrier and signal components are then separated in the detector or demodulator, $e_7(t)$. The carrier component is used only to enable the signal to be conveyed through the transmission medium and, after detection or demodulation, is discarded in the receiver. The signal component, on the other hand, represents the intelligence to be conveyed. It is usually amplified at signal frequencies and is then converted into acoustic energy, $p_2(t)$, by the loud-speaker. If the system is properly designed, the reproduced signal, $p_2(t)$, will produce essentially the same sensations as that of which the original signal, $p_1(t)$, is capable.

All the operations from $p_1(t)$ to $p_2(t)$ in Fig. 2 come within the province of the radio engineer. Since most radio engineering design problems are electrical in character and the initial and final steps involve physiological and psychological matters in which electrical engineers are not usually too proficient, it is convenient to separate the purely electrical operations of the system from those in which physiological responses and the conversion of energy from a sense-perceptive form into electrical form (or vice versa) are involved. The purely electrical portions of the radio system extend from $e_1(t)$ to $e_8(t)$; the initial and final steps are concerned with physiological responses.

PHYSIOLOGICAL RESPONSES AND THE COMMUNICATION SYSTEM

7. Role of Sensation Responses. The type or character of the signal to be transmitted through a radio transmission path depends upon the kind of intelligence to be conveyed. It also depends upon that sense which is to be stimulated at the receiving end of the system. For example, if only sound signals are to be transmitted, the appeal is made to the sense of hearing; the energy-converting devices at the transmitting and receiving ends must be designed to convert acoustic energy into electrical energy or vice versa. In television systems both the visual and aural senses are appealed to, and hence both electrovisual and electroacoustic energy-converting mechanisms are required at both ends of the system. A study of energy-converting devices, therefore, has its proper place in radio engineering.

The energy conversions in an electrical communication system require a pair of inversely related devices to convert the sense-perceptive energy to electrical energy at the transmitting end and from electrical energy to sense-perceptive energy at the receiving end. The satisfactory design of energy-converting devices involves three factors:

1. Physical means for converting sensation-stimulating energy into electrical form, and vice versa. This problem is one of utilizing the forces of nature and falls within the province of those trained in engineering or the physical sciences.

2. Means for determining the performance of the energy-converting devices in terms of physical energy as well as in terms of the sensations produced by their proper use. The criterion of ideal transmission, already given, may be employed to guide design. It may be technically impossible or economically unsound to construct an ideal (or even a completely distortionless) communication system. But by being acquainted with the conditions for ideal communication, the engineer is in a position to recognize the limitations beyond which engineering practice cannot proceed.

3. Knowledge of the relations between sensations or subjective responses and physical stimuli or certain physiological characteristics of the human being. It is particularly helpful to know quantitatively the important characteristics of speech, hearing, and vision, since this knowledge is of direct aid in the design of significant components of the radio system.

The argument is often advanced that, since the communication system ultimately appeals to one's senses, its behavior should be judged by the subjective sensations that it produces rather than by quantitative objective measurements of its inherent physical characteristics. Such argument fails to take account of the deviations from normal sense perceptions that are present in a significant proportion of the population; it also overlooks the important consideration that what looks or sounds well to one person may not be completely acceptable or appear normal to another. The exclusively subjective approach also fails to make possible any quantitative measure by which the performance of the system may be judged and specified and by which definite improvements can be made in the system even though they may be, individually, so small as to produce hardly perceptible changes in sense perception.

If the communication system is to be designed and built for use by normal individuals, it is desirable to know, quantitatively, the normal human responses from a statistically significant group of persons. The data obtained from a study of the subjective sensations and characteristics of a statistically adequate group of persons may be used as the basis for the design of communication equipment for any type of service required for normal persons.

8. Physiological Characteristics of Sound. From the physicist's point of view, sound is a longitudinal wave motion of alternate condensations and rarefactions in the transmission medium, which is nearly always air. From the psychological point of view, sound is a sensation produced by the organs of hearing when actuated by the energy of the condensations and rarefactions of wave motion. It is necessary to keep clearly in mind which concept is being used.

The Weber-Fechner law of psychology states that the increase of a stimulus necessary to produce a just discernible change in the resulting sensation bears a constant ratio to the total stimulus. The sensation produced increases or decreases arithmetically as the stimulus, measured in terms of energy or power density, is increased or decreased geometrically. The magnitude of the sensation produced is, therefore, proportional to the logarithm of the energy or power density of the stimulus. Although this law is much too simple a statement to account precisely for observed effects, in a general way it does represent the relation between subjective sensation and physical stimulus, at least for aural and visual effects; it is certainly more nearly correct than a linear correlation between the stimulus and the sensation. It is, in fact, a sufficiently good approximation to provide a sound psychological basis for the use of logarithmic units such as the neper, bel, decibel (db), and volume unit in those numerous cases for which the numerical designations of measurements are intended to convey a reasonably accurate correlation between the energy of the stimulus and the magnitude of the resulting sensation.

The sound intensity of a sound field, in a specified direction at a point, is the sound energy transmitted, per unit of time, in the specified direction through a unit area that is perpendicular to the specified direction at the point. The customary units of measuring sound intensity are the dyne per square centimeter (which has also been given

the name "bar") and the watt per square centimeter. The sound intensity, is therefore, a physical stimulus giving rise to the sensation of loudness.

9. **Loudness Level.** In acoustics it is much more customary to measure the loudness level than to measure the loudness, which is the magnitude of the sensation. Loudness level is the intensity level of an equally loud reference tone, which is usually taken to be 1,000 cps. Loudness level is an experimentally determined quantity and is measured in phons above a specified reference level. For loudness comparisons the reference level is 10^{-16} watt per sq cm, or 0.000204 bar.

The average characteristics relating frequency and loudness level for normal human ears are shown in Fig. 3, in which the frequency scale is logarithmically spaced. The ordinate scale at the right is logarithmically spaced and marked in absolute values

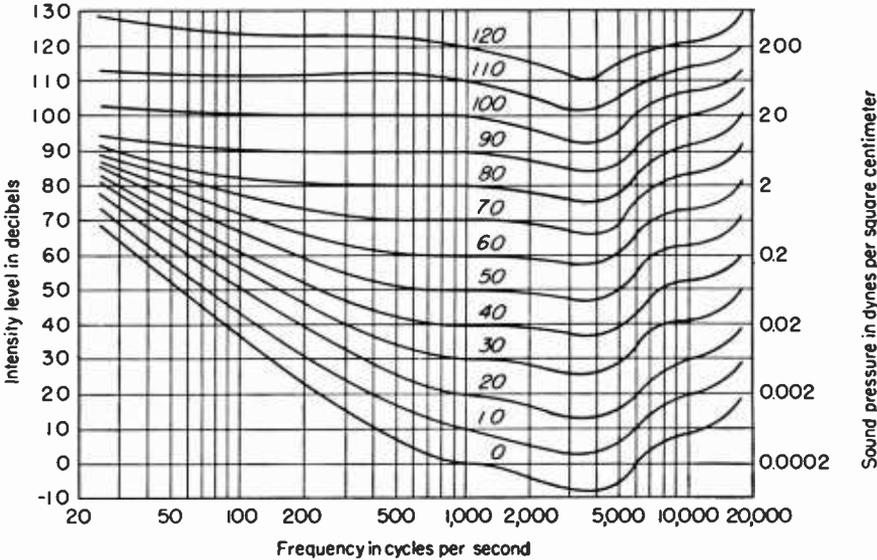


FIG. 3. Contours of equal loudness for normal human ears. Numbers on the curves indicate the loudness levels. The intensity level is given in decibels and the corresponding sound pressure in dynes per square centimeter.

of sound-pressure density, whereas the intensity-level ordinate scale at the left is linear in logarithmic units of decibels. The lowest curve, 0, of Fig. 3 represents the threshold of hearing below which no sound is heard by the normal human ear. The top curve represents the threshold of pain above which a sensation of pain rather than sound is produced.

Two characteristics that are important to the communication engineer are evident from Fig. 3. The normal human ear is responsive to frequencies of from about 20 to about 20,000 cps, covering a range of 10 octaves. Moreover the sensitivity of the ear varies considerably with frequency, being most sensitive, as judged by measurements of loudness level, for frequencies of from 2,000 to 5,000 cps. The curves of Fig. 3 also illustrate the loss of hearing at low intensity levels for high and especially for low frequencies. The maximum range of intensity to which the ear responds is about 140 db, representing a power ratio of 10^{14} :1.

10. **Frequency and Power Ranges for Practical Communication.** From the curves of Fig. 3 it may be determined that distortionless transmission of speech requires that the electrical system (including the electroacoustic transducers) must be capable of transmitting frequencies from 20 cps or less to 20,000 cps or more and that the system

must be capable of producing a range of acoustic power of 10^{14} to 1. Few electrical communication systems respond fully over this extended frequency range, and none of them is capable of operating over the full range of power ratios to which the normal human ear responds.

Fortunately it is not necessary to reproduce the complete range of frequencies and loudness levels to obtain satisfactory and understandable (even enjoyable) reproduction of speech and music. Estimates vary as to the frequency range and intensity levels that are desirable in communication, particularly since economic factors must also be taken into account because range extension (of both frequency and intensity level) is usually accompanied by increased cost. As shown by the single cross-hatched area of Fig. 4, little real improvement in the reproduction of music is obtained

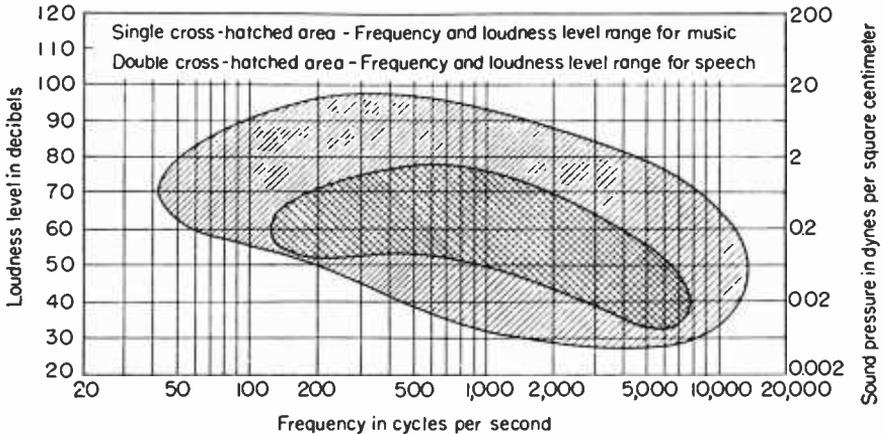


FIG. 4. Frequency and power ranges for satisfactory reproduction of speech and music.

by extending the frequency range below 40 or above about 15,000 cps or by extending the intensity levels below 28 or above 97 db. On the other hand, for persons of normal hearing, if the system reproduces frequencies from 130 to 5,000 cps and loudness levels of from 35 to 78 db, practically no gain justified by the additional cost will be achieved by exceeding these limits in the reproduction of speech. Indeed, intelligible speech can be reproduced with a frequency band from 250 to 2,800 cps, and for the reproduction of music many broadcast receivers are so operated that their frequency band is limited to between about 100 and 3,000 cps.

11. Sensation-Stimulus Relation. From the loudness-level curves of Fig. 3 it is possible to derive another set of relations in which loudness level is plotted against the sound pressure, as in Fig. 5. Such a relation shows directly the subjective sensation as a function of the stimulus within the framework of the method of measurement. If the Weber-Fechner law were rigorously true, the curves of Fig. 5 would be straight lines. The numbers on the individual curves designate the frequency in cps for which the curve applies.

The subjective response of the ear is not a linear function of the amplitude of the stimulus if a sufficient range of the latter is taken into account. Measurements by Stevens and Newman¹ show that although the stimulus-sensation response is linear for small values of sound-pressure intensity, amplitude distortion is produced within the ear itself for large amplitudes of sound.

12. Pitch and Frequency. Further experiments by Stevens and Newman have shown that, for the technique they have employed, the subjective musical pitch is

¹ Stevens, S. S. and E. B. Newman, On the Nature of Aural Harmonics, *Proc. Natl. Acad. Sci. U.S.*, **22** (11) 668, November, 1936.

not quite the same as the physical pitch measured in cps. The results of these investigations cannot be compactly summarized; original sources should be consulted.¹

13. Phase Shift. Conclusive evidence has not been obtained to indicate that the ear can evaluate phase shift in the components of a complex sound. The response of the human ear to complex waves appears to depend upon the magnitude of the harmonic components and not upon the wave shape which, for a specified amplitude-frequency spectrum, is determined by the phase shift of the harmonics compared to that of the fundamental. It is for this reason that the phase-shift characteristics of

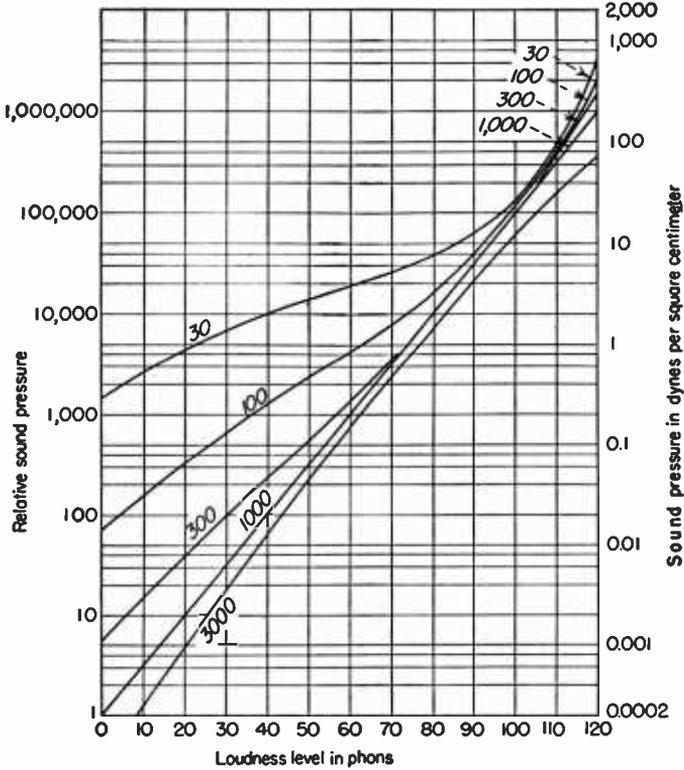


FIG. 5. Relative loudness (subjective sensation) plotted against sound pressure (physical stimulus) for five different audio frequencies. Deviations from linearity indicate discrepancies from the Weber-Fechner relation.

audio-frequency (a-f) systems are often neglected. Although the ear itself cannot distinguish the phase relations of the harmonics, there are subsidiary (although usually minor) matters that are observed as a result of phase shift. In a complex wave the phase relations of the harmonics may influence the subjective sensations of a complex tone. Provided the fundamental is less than 100 cps a phase shift of a harmonic may alter the subjective loudness of that harmonic and may also produce a noticeable difference in the quality of the complex tone.

Only a few of the more important physiological factors of hearing can be included here, but the references will serve as a guide to more detailed treatises.

¹ Stevens, S. S., and H. Davis, "Hearing, Its Psychology and Physiology," Wiley, 1938.

References

Physical Aspects of Sound:

- Barton, E. H.: "A Textbook of Sound," Macmillan, 1908.
 Crandall, I.: "Vibrating Systems and Sound," Van Nostrand, 1926.
 Fletcher, H.: "Physical Characteristics of Speech and Music," *Bell System Tech. J.*, 10 (3), 349, July, 1931.
 ———: "Speech and Hearing," Van Nostrand, 1929.
 Grimsehl, E.: "A Textbook of Physics," Blackie, 1933.
 Jeans, J. H.: "Science and Music," Macmillan, 1937.
 Jones, A. T.: "Sound, A Textbook," Van Nostrand, 1937.
Journal of the Acoustical Society of America.
 Lamb, H.: "Dynamical Theory of Sound," Arnold, London, 1931.
 Miller, D. C.: "The Science of Musical Sounds," Macmillan, 1937.
 Morse, P. M.: "Vibration and Sound," McGraw-Hill, 1948.
 Rayleigh, J. W. S.: "Theory of Sound," Macmillan 1929; Dover, 1945.
 Tyndall, J.: "On Sound," Longmans, 1896.
 Watson, F. R.: "Sound," Wiley, 1935
 Wood, A. B.: "Sound," Macmillan, 1944.

Physiological Aspects of Sound:

- Bartholemew, W. T.: "Acoustics of Music," Prentice-Hall, 1942.
 Beasley, W. C.: Characteristic Distributions of Impaired Hearing in the Population of the United States, *J. Acous. Soc. Am.*, 12 114, July, 1940.
 Best, E. H., and N. B. Taylor: "Physiological Basis of Medical Practice," Section 9, Williams & Wilkins, 1945.
 Crandall, I. B.: Speech; Dynamical Study of Vowel Sounds, *Bell System Tech. J.*, 6 (1), 100, January, 1927.
 ———, and D. Mackenzie: Analysis of Energy Distribution in Speech, *Bell System Tech. J.*, 1 (1), 116, July, 1922.
 Douglas, S.: "The Voice; Its Production and Reproduction," Pitman, 1933.
 Fletcher, H.: The Nature of Speech and Its Interpretation, *Bell System Tech. J.*, 1 (1), 129, July, 1922.
 ———: Physical Measurements of Audition and Their Bearing on the Theory of Hearing, *Bell System Tech. J.*, 2 (4), 145, October, 1923.
 ———: Pitch, Loudness and Quality of Musical Tones, *Am. J. Phys.*, 14 (4), 215, August-September, 1946.
 ———, and W. A. Munson: Loudness; Its Definition Measurement and Calculation, *J. Acous. Soc. Am.*, 5 (2), 182, October, 1933; *Bell System Tech. J.*, 7 (4), 377, October, 1933.
 Licklider, J. C. R.: Effects of Amplitude Distortion on the Intelligibility of Speech, *J. Acous. Soc. Am.*, 18 (2), 429, October, 1946.
 Polyak, S. L., G. McHugh, and D. K. Judd: "The Human Ear in Anatomical Transparencies," McKenna, 1946.
 Sacia, C. F., and C. J. Beck: Power of Fundamental Speech Sounds, *Bell System Tech. J.*, 5 (3), 393, July, 1926.
 Sivian, L. J.: Speech Power and Its Measurement, *Bell System Tech. J.*, 5 (4), 646, October, 1929.
 Snow, W. B.: Audible Frequency Ranges of Music, Speech and Noise, *J. Acous. Soc. Am.*, 3 (1), Part 1, 155, July, 1931.
 ———, H. C. Montgomery, and M. B. Gardner: Results of the World's Fair Hearing Tests, *Bell System Tech. J.*, 19 (4), 533, October, 1940.
 Stevens, S. S., and H. Davis: "Hearing, Its Psychology and Physiology," Wiley, 1938.
 ———, and E. B. Newman: On the Nature of Aural Harmonics, *Proc. Natl. Acad. Sci.*, 22 (11), 668, November, 1936.
 Steinberg, J. C.: Articulation Testing and Methods, *Bell System Tech. J.*, 8 (4), 806, October, 1929.
 Wiley, L. E.: "Function of the Brain in Audition," Philadelphia, 1932.

Acoustics:

- Davis, A. H.: "Modern Acoustics," Macmillan, 1927.
 ———, and G. W. C. Kaye: Acoustics, Bell, 1927.
 ———, and C. J. Morreau: "Reduction of Noise in Buildings," H. M. Stationery Office, 1939.
 Hughes, L. E. C.: "Elements of Engineering Acoustics," Benn, 1933.
 Knudsen, V. O.: "Architectural Acoustics," Wiley, 1940.
 ———: "Modern Acoustics and Culture," University of California Press, 1937.
 McLachlan, W. N.: "The New Acoustics," Oxford, 1936.
 Morse, P. A., and R. H. Bolt: "Sound Waves in Rooms," American Institute of Physics, 1945.
 Motion Picture Arts and Sciences, Academy, "Motion Picture Sound Engineering," Van Nostrand, 1938
 Olson, H. F.: "Elements of Acoustical Engineering," Van Nostrand, 1940.
 ———, and Frank Massa, "Applied Acoustics," Blakiston, 1939.
 "Research in Sound in the Theater," Stevens Institute of Technology, 1941-1942.
 Sabine, P. E.: "Acoustics and Architecture," McGraw-Hill, 1932.
 ———: "Collected Papers on Acoustics," Harvard University Press, 1922.
 Sound Insulation and Acoustics, *Dept. Sci. Ind. Research*, H. M. Stationery Office, 1945.

"Standards on Electroacoustics," IRE, 1938.
 Stewart, G. W.: "Introductory Acoustics," Van Nostrand, 1933.
 ———, and R. B. Lindsay: "Acoustics," Van Nostrand, 1930.
 Watson, F. R.: "Acoustics of Buildings," Wiley, 1930.
 West, W.: "Acoustical Engineering," Pitman, 1932.

14. Some Important Characteristics of Vision. The introduction of picture transmission and television has made it imperative that the communication engineer have at least an elementary acquaintance with some of the more important normal visual characteristics of humans. Human visual characteristics have not been associated with problems in communication for so long a time as aural characteristics, but in either case engineering design is facilitated by a knowledge of normal human characteristics and behavior.

15. Visual Sensitivity. The subjective sensation of brightness is, roughly, proportional to the logarithm of the stimulus from a white light source, as indicated in Fig. 6. At relatively low and also at very high values of field brightness there are deviations from the Weber-Fechner relation.

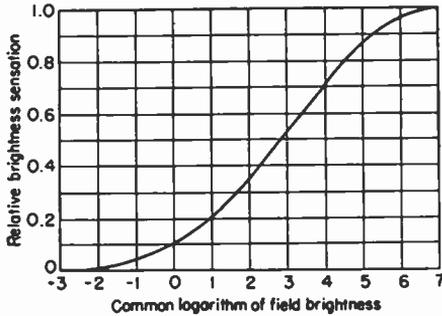


FIG. 6. Relation between subjective response to white light having field brightness expressed as the logarithm of intensity in candles per square meter.

The sensitivity of the eye to light is not uniform throughout the visible spectrum of from 400 to 700 $m\mu$ but varies with the wavelength of the radiation. To a certain extent the eye sensitivity varies also with the intensity of the light. For values of light intensity normally encountered, the visibility (or color sensitivity) curve of the normal human eye has its peak at 554 $m\mu$, as shown in Fig. 7. At very low light levels the visual mechanism differs from that which operates at high values, and the visual curve is shifted about 40 $m\mu$ toward the lower wavelengths with the peak at 515 $m\mu$. This shift is known as the "Purkinje effect."

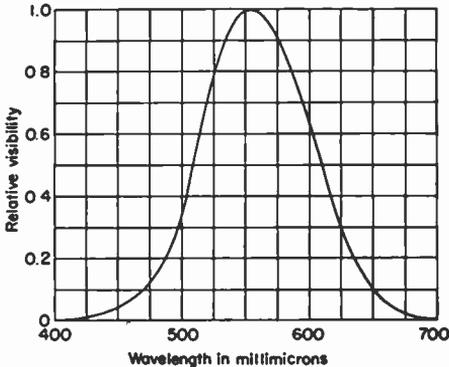


FIG. 7. Visibility curve of normal human eye for scopic vision (normal levels of luminosity intensity).

indicates the degree to which visual detail should be provided if the images reproduced by the electrical system are to be essentially like the original scene.

17. Persistence of Vision and Flicker. Experiments show that a visual image may be fully perceived by a viewer about 0.01 sec after it has been projected onto a screen. The visual impression does not immediately vanish upon being removed from the screen suddenly but persists as an afterimage for about 0.05 sec. The persistence of vision is employed to good advantage in motion pictures and television

16. Visual Angle. The minimum angular separation between two black dots on a white field that can just be resolved by persons with normal vision is called the "minimum perceptible visual angle." It is measured in minutes of arc, and the numbers expressing visual acuity increase as vision improves. Visual acuity is not a constant but depends upon the intensity of the light by which tests are made, as shown in Fig. 8. Again there is an approximately linear response to a logarithmic stimulus. The normal visual acuity

to create the illusion of continuous motion by exhibiting, in rapid succession, a series of images that are similar to one another but individually slightly different in those regions in which motion occurs. If 20 or more such images are viewed per second and if each successive image can be reconstructed before the visual impression of the preceding one has been lost, the sensation produced is one of apparent continuous motion, and the images are reasonably free from flicker. Measurements of persistence of vision are aided by determining the frequency of repetition at which alternate black-white images fail to produce flicker. The flicker frequency, as shown in Fig. 9, is not constant but depends upon the logarithm of the field brightness.

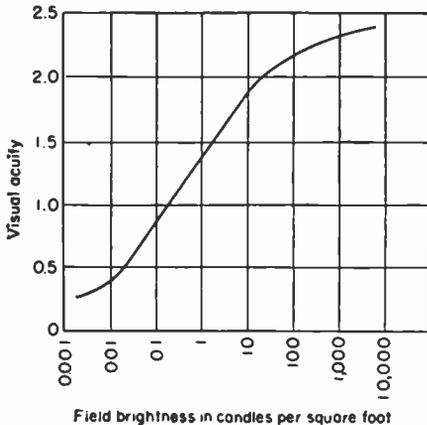


Fig. 8. Visual acuity as function of field brightness.

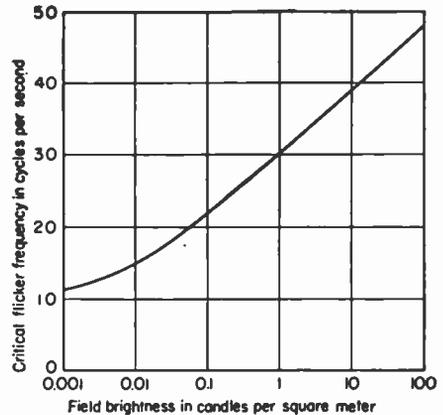


Fig. 9. Critical frequency, in flashes per second, plotted as a function of field brightness.

The experimental fact that any color stimulus can be subjectively matched by a mixture of three arbitrarily chosen primary colors provides the basis by which visual images may be produced in color.

Although a fairly extensive treatment of human visual characteristics is of interest to the engineer designing or using electrovisual devices, only a few of the more important characteristics can be mentioned here. The reader is referred to the more specialized literature for more extensive treatment.

References

Physical Aspects of Light and Optics:

- Bassett, A. B.: "Treatise on Physical Optics," 1892.
 Drude, P.: "Theory of Optics," Longmans, 1902. Later printing, 1929.
 Hardy, A. C., and F. H. Perrin: "Principles of Optics," McGraw-Hill, 1932.
 Houston, R. A.: "A Treatise on Light," 6th ed., Longmans, 1930.
 Moon, P.: "Scientific Basis of Illuminating Engineering," McGraw-Hill, 1936. (Contains many references.)
 Preston, T.: "Theory of Light," Macmillan, 1901.
 Southall, J. P. C.: "Introduction to Physical Optics," Oxford, 1937.
 Valasek, J.: "Elements of Optics," McGraw-Hill, 1928.
 Wood, R. W.: "Physical Optics," Macmillan, 1902. New ed., 1929.

Physiological Aspects of Light and Vision:

- Adams, D.: "Dark Adaptation; A Review of Literature," Oxford, 1929.
 Best, C. H., and N. B. Taylor: "Physiological Basis of Medical Practice," Williams & Wilkins, 1945.
 Cady, F. E., and H. B. Dates: "Illumination Engineering," Wiley, 1928.
 Ferree, C. E., and G. Rand: "Studies in Physiological Optics," Vols. I and II, Wilmer Ophthalmological Institute, 1934.
 Fulton, J. F., P. M. Hoff, and H. T. Perkins: "Bibliography of Visual Literature, 1939-1944," Office of Scientific Research and Development, 1945.
 von Helmholtz, H.: "Physiological Optics," Optical Society of America, 1924.

Houston, R. A.: "Vision and Color Vision," Longmans, 1932.
 Klüver, H.: "Visual Mechanism," Jacques Cattell Press, 1942.
 Kronfeld, P. C., G. McHugh, and S. L. Polyak, "The Human Eye in Anatomical Transparencies," Bausch and Lomb, 1943.
 Lickley, J. D.: "The Nervous System," Longmans, 1919.
 Luckiesh, M.: "Light, Vision and Seeing," Van Nostrand, 1944.
 ——— and F. K. Moos: "The Human Seeing Machine," *J. Franklin Inst.*, 215 (6), 629, June, 1933.
 ——— and ———: "The Science of Seeing," Van Nostrand, 1937.
 ——— and ———: "Seeing—A Partnership of Lighting and Vision," Williams & Wilkins, 1931.
 Lythgoe, R. J.: "Measurement of Visual Acuity," Oxford, 1932.
 ——— and K. Tansley, "Adaptation of the Eye," University of London Press, 1929.
 Moon, P.: "Scientific Basis of Illuminating Engineering," McGraw-Hill, 1936.
 Nutting, P. G. (Chairman): Report of the Standards Committee on Visual Sensitometry, *J. Optical Soc. Am.*, 4 (2), 55, March, 1920.
 Parsons, J. H.: "Introduction to the Study of Color Vision," Macmillan, 1934.
 Walls, G. L.: "The Vertebrate Eye," Cranbrook Institute of Sciences, 1942.
 Weston, H. C.: "Relation between Illumination and Visual Efficiency," H. M. Stationery Office, 1945.
 Woodworth, R. S.: "Psychology," Holt, 1935.

18. Logarithmic Measurements. In communication engineering advantage is taken of the approximate experimental relation that the sensation is a logarithmic function of the stimulus in a number of useful units of measurement in which power ratios are expressed logarithmically. These logarithmic units provide not only a useful measure of the relation between subjective sensation and physical stimulus, but also a convenience in computing certain important properties of the communication system. A number of logarithmic units, particularly the decibel, have been found to represent such useful concepts that they have been improperly (and sometimes ridiculously) extended to fields in which they do not apply.

The *bel*, named in honor of Alexander Graham Bell, is defined as the common logarithm of the ratio of two powers. Thus, if P_1 and P_2 represent the powers at two different levels, the number of bels, N_b , expressing the change of power from P_1 to P_2 is

$$\text{Number of bels} = N_b = \log_{10} \frac{P_2}{P_1}$$

If P_2 is greater than P_1 , N_b will be a positive number (representing a gain in power); if $P_2 = P_1$, then N_b will be zero; and if P_2 is less than P_1 , N_b will be a negative number representing a loss in power.

The bel is a unit larger than is convenient to use in practical work, so that a smaller unit, the *decibel* (db), has been established for engineering work. The magnitude of a decibel (a decimal submultiple of the bel) is one-tenth that of the bel, so that for equal power ratios, ten times as many decibels as bels are required to express the power ratio. Therefore, the number of decibels required to express the change in power level from P_1 to P_2 is

$$\text{Number of decibels} = N_{db} = 10 \log_{10} \frac{P_2}{P_1}$$

The decibel is of such size that a change of 1 db in acoustic power can just be detected.

Another unit, often used in theoretical work because of the mathematical convenience in dealing with exponents in the natural system is the *neper*, named in honor of John Napier, originator of the natural system of logarithms. The neper is defined as one-half the natural logarithm of the ratio of the two powers. If P_1 and P_2 represent the two power levels, the number of nepers expressing the change in power level from P_1 to P_2 is

$$\text{Number of nepers} = N_n = \frac{1}{2} \log_e \frac{P_2}{P_1}$$

where e is the base of the natural system of logarithms and has the approximate value $e = 2.7182818285 \dots$

A definite relationship exists between the power ratios expressed in these three systems. Thus,

$N_n = (\log_e 10)N_b = 1.151292546 \dots N_b$	or	1 bel = 1.151 nepers
$= \frac{1}{2} \log_{10} (\log_e 10) N_{db} = 0.1151292546 \dots N_{db}$	or	1 decibel = 0.1151 neper
$N_{db} = 10N_b$	or	1 bel = 10 decibels
$= 20(\log_{10} e)N_n = 8.68588964 \dots N_n$	or	1 neper = 8.686 decibels
$N_b = 0.1N_{db}$	or	1 decibel = 0.1 bel
$= 2(\log_{10} e)N_n = 0.868588964 \dots N_n$	or	1 neper = 0.8686 bel

These terms express only a power ratio, and the number of units is *not* an expression for an absolute value of power. By assigning the value of zero (0) to some selected reference level, it is possible to ascertain an absolute value of power for any numerical expression in any of the three systems. Values of 6, 10, and 100 mw have been used as the reference level at different times by different groups, so that the reference level should always be specified as well as the numerical units for a given power ratio. The change in power level is always a numeric expressing an increase or decrease of power level (depending upon whether P_2/P_1 is greater or less than unity, respectively); it is never properly expressed as the absolute value of power.

Power may be expressed in terms of voltage or current and resistance, since $P = I^2R = E^2/R$, where I is the current through the resistance, R , and E is the voltage across the same resistance. Hence it is possible for any of the above expressions to take the general form (but not the exact value)

$$N = \log \frac{P_2}{P_1} = \log \frac{I_2^2 R_2}{I_1^2 R_1} = \log \frac{E_2^2 R_1}{E_1^2 R_2}$$

When a change of power level is expressed in terms of resistance and either current or voltage, it is generally necessary to specify the values of resistance, R_1 and R_2 , as well as the values of the voltages or currents for the two power levels. Only in the special case for which $R_1 = R_2$ does neglect of the values of the resistance introduce no error. Yet it is a common if regrettable practice to overlook the values of the resistance.

The *volume unit* (vu) is defined as ten times the common logarithm of the power ratio, P_2/P_1 , where the reference level, P_1 , is selected as 1 mw (0.001 watt). If P_2 is the power level measured in watts, the number of volume units expressed by the ratio P_2/P_1 is

$$\text{Number of volume units} = N_{vu} = 10 \log_{10} \frac{P_2}{0.001} = 10 \log_{10} 1,000P_2 = 30 \log_{10} P_2$$

Since the reference level is specified in the definition of the volume unit, this term can be properly used to express either a change in power level or an absolute value of power.

The *phon* is a unit of loudness level. The loudness level of a sound, in phons, is numerically equal to the intensity level (in decibels) of a pure tone having a frequency of 1,000 cps, which is judged by the listener to be of equivalent loudness. In establishing the phon unit, zero reference level is taken as the average limit of audibility at 1,000 cps; this has been standardized as 10^{-16} watt per sq cm or 0.000204 dyne per sq cm. Since the phon is a unit representing a subjective sensation, its value is directly affected by the average characteristics of the normal human ear. The number of phons representing a given sound intensity level is, in general, different for each frequency and varies in a manner depending upon the sensitivity of the normal human ear as shown by the loudness level contours of Fig. 3.

References

- Chinn, H. A., D. K. Gannett, and R. M. Morris: A New Standard Volume Indicator and Reference Level, *Proc. IRE*, **28** (1), 1, January, 1940.
 Gray, C. H. G., and W. H. Martin: Master Reference System for Telephone Transmission, *Bell System Tech. J.*, **8** (3), 536, July, 1929.
 Hilliard, J. K.: Definition of Standard Reference System, *Electronics*, November, 1931.
 Martin, W. H.: Decibel—The Name for the Transmission Unit, *Bell System Tech. J.*, **8** (1), 1, January, 1929.

ANALYTICAL APPROACH TO ELECTRICAL COMMUNICATION PROBLEMS

19. General Method. In general, the method of attacking and solving problems in electrical communication usually involves the following steps.

1. The differential equation expressing the dynamic equilibrium of the system is set up; with initial and boundary conditions specified.
2. The function satisfying that equation is determined by a process of integration which yields the formal solution to the problem. The solution thus found is the correct one, although it may be more general than is required.
3. The formal solution must then be subjected to the initial conditions of the system as well as to the boundary conditions of time and space to obtain the solution for the particular problem under consideration.

If they are to be useful for communication purposes, the differential equations of dynamic equilibrium, as well as the formal solutions of these equations, must be expressible in terms of the intelligence-bearing signals. It becomes necessary, therefore, to obtain analytical expressions for the representation of intelligence. Such expressions should be as simple as possible for mathematical manipulation consistent with the requirement that they shall be sufficiently rigorous and useful for engineering and scientific work.

The great number of signals encountered in communication systems may be represented by periodic, nonperiodic, and transcendental functions. For communication purposes the functions expressing signals must be single-valued functions of time. Strictly speaking, the process of communicating intelligence is carried on without repetition so that nonperiodic functions, expressed in terms of the Fourier integral, provide the most accurate approach. Unfamiliarity with and the difficulty of handling Fourier integrals on the part of engineers have limited the use of nonperiodic functions in radio engineering analysis, however.

A great deal of information may be determined about the behavior of the communication system by dealing with the mathematically much simpler periodic functions expressing the steady-state dynamic equilibrium of the system. Because of the amount of useful information derivable from steady-state analysis and the comparative ease with which periodic functions may be manipulated mathematically, the analysis of communication systems has usually been made through the use of periodic functions of time.

20. Simple Harmonic Functions. The simplest type of periodic function is that which varies harmonically with respect to time. Such a function may be expressed as

$$y = f(t) = E \sin(\omega t + \phi) = E \cos\left(\omega t + \phi + \frac{\pi}{2}\right)$$

where E is the amplitude or maximum value of the function, $\omega = 2\pi f$ is the angular frequency in radians per second if f is the cyclic frequency in cps, and ϕ is the phase angle relating the harmonic function with that point of the cycle from which time is measured. Except for matters involving phase shift, statements applicable to the sine function are also generally true for the cosine function.

The harmonic functions are mathematically simple to manipulate and represent quite accurately many kinds of functions or wave forms encountered in communication systems. Moreover, they provide the basis for dealing with much more complex periodic functions, expressible by means of the Fourier series, since any periodic, recurrent function may be analyzed into, or synthesized from, a series of simple harmonic functions.

Harmonic functions involve three constants: the amplitude, E , the angular frequency, ω (or the cyclic frequency, f), and the phase displacement, ϕ . When these three constants are known, the harmonic function (either the sine or the cosine) is completely specified. Harmonic functions may be illustrated graphically by means of the familiar sine-wave curve of Fig. 10. Such a representation has the advantage of illustrating, pictorially, the manner in which the function varies with time. From

such a graph, the three characteristics (E , ω , and ϕ) may be determined. An amplitude-time curve such as that of Fig. 10 is a satisfactory method of representation so long as a single harmonic function or, at most, only a few such functions are to be dealt with, but it does not lend itself well to those numerous cases in which many harmonic functions may be required simultaneously to represent the wave forms encountered in practice.

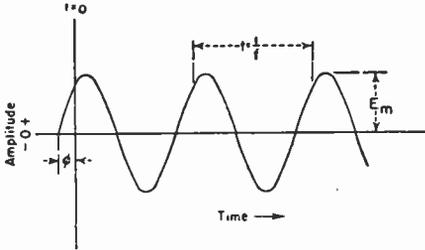


FIG. 10. Harmonic function plotted explicitly against time. E_m is the amplitude of the function, ϕ is the phase displacement, and t is the time required for a complete cycle. The angular frequency in radians per second is $\omega = 2\pi f$, where f is the cyclic frequency in cycles per second.

formance of the communication system on a frequency basis.

The great advantage of the frequency spectra plots become evident as it becomes necessary to employ more and more functions to express a complex, periodic wave form. The center diagram of Fig. 11 is called the "amplitude-frequency" spectrum and the lower diagram is called the "phase-frequency" spectrum of the harmonic function. The amplitude and phase plots for the harmonic variation in the upper curve are also indicated. Usually the cyclic frequency, f , in cps, is used as the abscissa instead of the radian frequency, ω , although either may be employed, since $\omega = 2\pi f$.

If the function representing the signal or intelligence is more complicated than can be accurately expressed by a single harmonic function, but if it is, nevertheless, a periodic function of time, it may be considered to be made up of the superposition of a number of harmonic functions, each of which is characterized by its individual values of E , ω , and ϕ . These three characteristic constants for each harmonic component may then be plotted as the amplitude-frequency and phase-frequency spectra. The spectra for a rectangular wave are shown in Fig. 12 for several of the lowest frequency components of such a periodic wave form.

22. Analytical Expressions for Simple Harmonic Functions. Simple harmonic functions may be regarded as being generated by the projection, on a suitable axis, of a vector of magnitude E , rotating with a constant angular velocity of ω radians per sec, with an initial phase displacement, ϕ , as shown in Fig. 13. The projection of the end of the rotating vector on the x , or real, axis yields the cosine function, whereas

21. Amplitude- and Phase-frequency Spectra. The simple harmonic function may also be plotted to show the amplitude-frequency and the phase-frequency relations as shown in Fig. 11. This pair of frequency plots (called a "pair of spectra") is capable of conveying much more information clearly than the pictorial plot, especially when many harmonic functions must be dealt with simultaneously. Moreover, the amplitude-frequency and phase-frequency spectra can be used to express the performance of the communication system on a frequency basis.

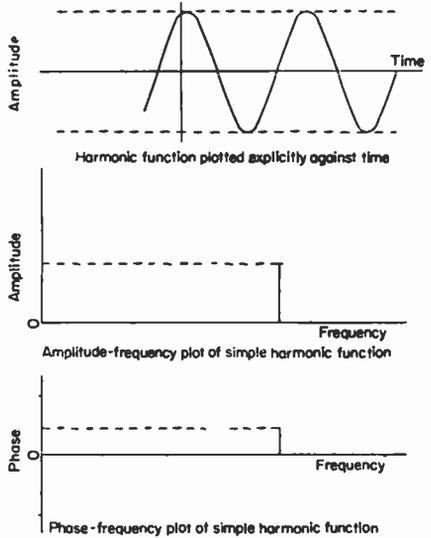


FIG. 11. Amplitude and phase spectra of harmonic function of time.

the projection of the rotating vector on the y , or quadrature (or imaginary), axis produces the sine function. At any instant of time the sine and the cosine functions are the rectangular components of the rotating vector as expressed by the vector sum

$$E = E \cos(\omega t + \phi) + jE \sin(\omega t + \phi)$$

where $j = \sqrt{-1}$ is an operator indicating rotation of 90 deg in the counterclockwise direction. The rotating vector may also be expressed in terms of its polar components, the amplitude and the phase, rather than in terms of its rectangular components. In polar coordinates,

$$E = |E|/\phi = \sqrt{E_{RE}^2 + E_{QU}^2}/\phi$$

where the real component is

$$E_{RE} = |E| \cos(\omega t + \phi),$$

the quadrature component is given by $E_{QU} = |E| \sin(\omega t + \phi)$, and the phase displacement is $\phi = \tan^{-1}(E_{QU}/E_{RE})$.

The rectangular form is more convenient when vector quantities are to be added or subtracted, whereas the polar form is more suitable when vectors are to be multiplied or divided.

23. Exponential Representation of Harmonic Functions. The rotating vector representing the harmonic function may be expressed in exponential form, which is particularly useful for mathematical operations involving differentiation and integration. Noting that the rotating operator, j , is defined as $j = \sqrt{-1}$ and is to be interpreted as indicating a counterclockwise rotation of 90 deg for the quantity on which it operates, successive applications of j yield $j^2 = -1$, or 180 deg rotation, $j^3 = -j$, or 270 deg rotation, and $j^4 = +1$, or 360 (or 0) deg counterclockwise rotation. Further successive powers of j merely result in repetition of this sequence.

NOTE: It is merely a coincidence that the fundamental period of the rectangular wave, upper plot, is drawn with the same spacing as that between harmonic components in the lower plots.

To show that a rotating vector of unit magnitude may be represented by means of exponentials, let the series expansion be written for e^{jx} , thus,

$$e^{jx} = 1 + \frac{(jx)}{1!} + \frac{(jx)^2}{2!} + \frac{(jx)^3}{3!} + \dots + \frac{(jx)^n}{n!} + \dots$$

If the algebraic value, $j = \sqrt{-1}$, is substituted into this series expansion and the series is then rewritten so that the real and quadrature (or j) terms are segregated, the result is

$$e^{jx} = \left(1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8!} - \dots\right) + j \left(x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \dots\right)$$

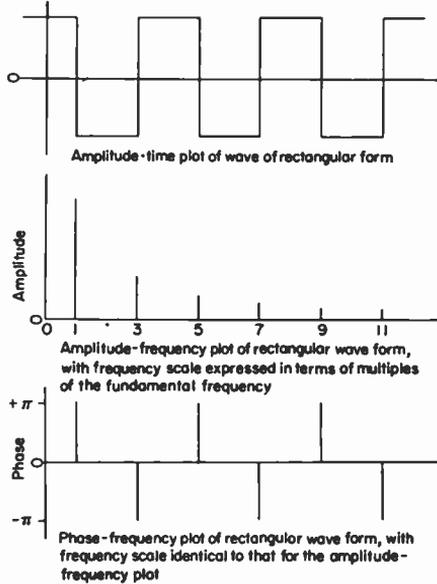


FIG. 12. Amplitude and phase spectra (center and bottom, respectively) of the steady-state square wave shown at the top. It is merely a coincidence that the spacing between harmonics in the spectra plots is the same as that between successive cycles of the square wave.

The series expansions for the harmonic functions of x are

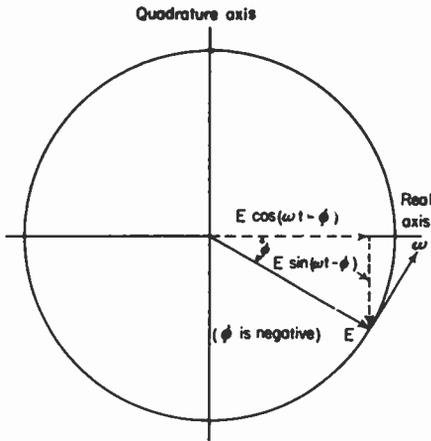


FIG. 13. Method of generating simple harmonic function by projecting, on a reference axis, the path described by the moving end of a vector of constant magnitude and rotating with constant angular velocity ω radians per sec, about a fixed end. The cosine function is generated by the projection on the real axis; the sine function by the projection on the quadrature axis as shown.

$$\cos x = 1 - \frac{x^2}{2!} + \frac{x^4}{4!} - \frac{x^6}{6!} + \frac{x^8}{8} - \dots$$

$$\sin x = x - \frac{x^3}{3!} + \frac{x^5}{5!} - \frac{x^7}{7!} + \frac{x^9}{9!} - \dots$$

A comparison of these two series with that for the exponential expansions shows that the rotating vector of unit length may be expressed in either of the forms of the equality

$$e^{ix} = \cos x + j \sin x$$

If the amplitude of the vector is E and if $x = (\omega t + \phi)$,

$$\begin{aligned} Ee^{i(\omega t + \phi)} &= Ee^{i\omega t}e^{i\phi} \\ &= E \sin(\omega t + \phi) + jE \cos(\omega t + \phi) \end{aligned}$$

In the same way, a vector of magnitude E rotating in the negative or clockwise direction with constant angular velocity, $-\omega$ radians per sec, may be expressed as

$$\begin{aligned} Ee^{i(-\omega t + \phi)} &= Ee^{-i\omega t}e^{i\phi} \\ &= E \cos(\omega t + \phi) - jE \sin(\omega t + \phi) \end{aligned}$$

In exponential terminology, the real component represents the cosine function according to the expression

$$E \cos(\omega t + \phi) = \text{Re}[Ee^{i(\omega t + \phi)}]$$

and the quadrature or imaginary component represents the sine function as given by

$$\begin{aligned} E \sin(\omega t + \phi) &= \text{Qu}[Ee^{i(\omega t + \phi)}] \\ &= \text{Im}[Ee^{i(\omega t + \phi)}] \end{aligned}$$

In engineering use the real component is usually employed and the Re notation is often dropped since it is understood that only the real component of the rotating vector is employed to express a harmonic variation of time.

Harmonic variations may also be expressed by two vectors of equal magnitude, rotating with equal but opposite angular velocities, ω and $-\omega$, and having the same initial phase displacement, ϕ , with respect to the reference axis when $t = 0$. As shown in Fig. 14, the cosine function may be expressed as one-half of the vector sum of two oppositely rotating vectors of equal magnitude according to the relation

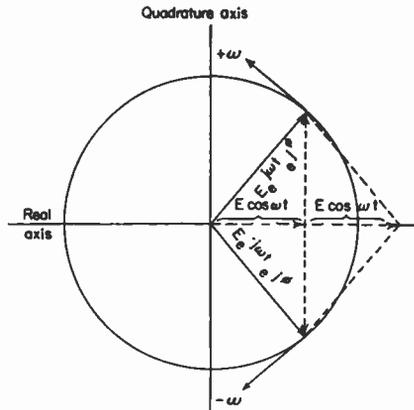


FIG. 14. Method of generating simple harmonic function by projecting on a reference axis the path described by the sum of two vectors of equal amplitude rotating in opposite directions with angular velocities of the same magnitude. In this diagram the initial phase displacement is $\phi = 0$.

$$e(t) = E \cos(\omega t + \phi) = \frac{1}{2}(Ee^{i\omega t}e^{i\phi} + Ee^{-i\omega t}e^{i\phi}) = \frac{1}{2}Ee^{i\phi}(e^{i\omega t} + e^{-i\omega t})$$

By forming the projection of the two vectors on the quadrature axis, the sine function likewise has amplitude equal to one-half that of the two equal vectors.

24. Fourier Analysis. Only exceedingly simple bits of intelligence may be adequately represented by the simple functions already discussed and, in general, much more complicated wave forms are required; hence, analytical expressions become necessary for expressing the more complex wave forms. According to Fourier's theorem, any periodic wave form may be expressed as the sum of a series of harmonic functions whose frequencies are integral multiples of the lowest or fundamental frequency. A constant term, which may be regarded as a harmonic function of zero (0) cyclic frequency, may also be required in some cases. The general Fourier series expansion for any particular periodic function may be written in the alternative forms

$$\begin{aligned}
 e(t) &= A_0 + \sum_{n=1}^{\infty} A_n \cos n\omega t + \sum_{n=1}^{\infty} B_n \sin n\omega t \\
 &= \sum_{n=-\infty}^{+\infty} C_n e^{jn\omega t}
 \end{aligned}$$

where $n = 0, 1, 2, 3, \dots$, is the order of the harmonic. For engineering use, the amplitude of successive terms in either and both forms of the Fourier series must decrease so that the series converges to a finite limit, but this condition is fulfilled for wave forms encountered in communication use.

For either of the above expressions to express properly a specified periodic function, it is necessary to know the numerical values of the amplitudes of the component harmonic functions of the Fourier series, *i.e.*, A_0, A_n and B_n , or C_n . To evaluate these coefficients, use is made of the following three properties of trigonometric integrals:¹

1. The average value of a single sine or cosine term, when taken over a complete period, is zero.
2. The average of the product of two sines, or of two cosines, or of a sine and a cosine, of commensurable but numerically unequal frequencies, taken over any complete period of the product, is zero.
3. The average value of the product of two sines or of two cosines of the same frequency taken over any complete period is one-half the cosine of the difference between the phase of the factors. For the product of a sine and a cosine, the average is one-half the sine of the angle by which the sine term leads the cosine term.

If $e(t)$ is an arbitrary periodic function of period T so that $e(t) = e(t + T)$, it is expressible as a Fourier series of harmonic functions. If the function, $e(t)$, is specified analytical or graphically, the Fourier coefficients may be determined from the relations

$$\begin{aligned}
 A_0 &= \frac{1}{T} \int_t^{t+T} e(t) dt \\
 A_n &= \frac{2}{T} \int_t^{t+T} e(t) \cos n\omega t dt \\
 B_n &= \frac{2}{T} \int_t^{t+T} e(t) \sin n\omega t dt \\
 C_n &= \frac{1}{T} \int_t^{t+T} e(t) e^{-jn\omega t} dt
 \end{aligned}$$

where $T = 2\pi/\omega$ is the fundamental period, or the period of the fundamental term. In the above expressions, C_n is a complex quantity related to A_n and B_n terms by the equations

$$C_0 = A_0 \quad C_{+n} = \frac{1}{2}(A_n - jB_n) \quad C_{-n} = \frac{1}{2}(A_n + jB_n)$$

¹ Franklin, P., "Differential Equations for Electrical Engineers," Wiley, 1933.

When the Fourier series represented by $e(t)$ is inserted in the integral expressions for the Fourier series coefficients, all terms, when integrated over a complete period, become zero except those for the n th term. Hence, for each value of n it is possible to determine the Fourier coefficients.

For most of the wave forms encountered in communication practice, the Fourier series requires an infinite number of terms to express completely and precisely the function $e(t)$. As a corollary, this condition implies that the response of the transmission system must have a frequency response extending to infinity if the periodic wave form is to be passed through it without change in wave form. Such an infinitely wide frequency band is not technically possible nor is too wide a band justified on economic grounds. For those wave forms which have practical value, the magnitude of the amplitudes of successively higher harmonics become progressively smaller. Therefore, no appreciable distortion of the wave shape will occur for signals passing through the transmission system, if some of the higher order harmonics of negligible amplitude are eliminated in the transmission system. In engineering design it is possible to establish an upper frequency limit for the transmission system such that the elimination of higher order terms of the Fourier series above the specified frequency limit will produce a negligible change in the wave form of the signal passing through the system. There is no straightforward and direct means for predetermining the highest frequency beyond which "negligible" change in wave form results. Engineering practice that is sometimes followed is to neglect the h-f components whose amplitudes are less than 10 per cent of that of the fundamental frequency, or to neglect harmonics beyond the tenth. The error in wave form thus introduced depends not only on what is understood by "negligible" change in wave form but also upon the shape of the wave represented by the function $e(t)$.

If $e(t)$ is a nonperiodic function, it cannot be expressed by a Fourier series as has been done above. It can be represented, however, by the limiting form that the Fourier series takes as the fundamental period, T , becomes infinitely great. The function $e(t)$ is then expressible by means of the Fourier integral, the solution of which yields a continuous pair of frequency spectra (the amplitude-frequency and the phase-frequency spectra) instead of the discontinuous pair of frequency spectra obtainable from the Fourier series in which each term represents a discrete frequency.¹

25. Average and Effective Values of Periodic Functions. If $e(t)$ is a periodic function of time with period T so that $e(t) = e(t + T)$, the average value of the function over a complete cycle of recurrent phenomena is

$$E_{av} = \frac{1}{T} \int_t^{t+T} e(t) dt$$

The average value of a periodic function that is symmetrical with respect to the zero axis throughout the cycle is zero. Over any integral number of complete cycles the average value of a sine or cosine function is, therefore, zero.

In speaking of the average value of a sine or cosine function, the electrical engineer refers to the average value of either the positive or the negative half of the cycle, the absolute value of which is

$$E_{av} = \frac{1}{\pi} \int_0^{\pi} E_m \sin \omega t d(\omega t) = \frac{2E_m}{\pi} \cong 0.6366198 \dots E_m$$

In many cases certain electrical effects are proportional to the square root of the average value of the square of the function, or to the rms or the effective value of the function. The effective value of the function, $e(t)$, is given by

¹ For treatment of the Fourier integral, see such texts as W. E. Byerly, "Fourier's Series and Spherical Harmonics," Ginn, 1893; J. R. Carson, "Electric Circuit Theory and the Operational Calculus," McGraw-Hill, 1928; T. v. Kármán and M. A. Biot, "Mathematical Methods in Engineering," McGraw Hill, 1940.

$$E_{rms} = \left[\frac{1}{T} \int_t^{t+T} [e(t)]^2 dt \right]^{1/2}$$

If $e(t) = E_m \sin \omega t$, the rms value of the function over a complete cycle or period is

$$E_{rms} = \left[\frac{1}{2\pi} \int_0^{2\pi} (E_m \sin \omega t)^2 d(\omega t) \right]^{1/2} = \frac{E_m}{\sqrt{2}} \cong 0.7071068 \dots E_m$$

By applying the equation for the effective or rms value to a periodic wave expressible by the Fourier series,

$$e(t) = E_0 + E_1 \sin(\omega t + \phi_1) + E_2 \sin(2\omega t + \phi_2) + \dots + E_n \sin(n\omega t + \phi_n)$$

it can be shown that the effective value of a periodic recurrent function is given by

$$e(t)_{rms} = [E_0^2 + \frac{1}{2}(E_1^2 + E_2^2 + \dots + E_n^2)]^{1/2}$$

26. Amplitude, Frequency, and Phase Characteristics of Electrical Communication

System. Having available the means for expressing intelligence analytically by Fourier series (or Fourier integrals) and also a means for determining the coefficients of the Fourier series (or the amplitude-frequency and phase-frequency spectra for the Fourier integral), it is now possible to establish conditions to be fulfilled by any transmission system in order that the signal passing through it may be free from distortion. The term "distortionless," when applied to a transmission system, signifies that the signal leaving the transmission system is to be identical in wave shape to that entering it; a delay in passing through the system is permissible since a finite (even though small) time is required for any physical phenomena to take place. For the Fourier-series method of expressing intelligence to hold, the transmission system must be linear; *i.e.*, the output response must be a linear response of the input function; hence, this condition imposes one limitation on the transmission medium if it is to be analyzed by the following method.

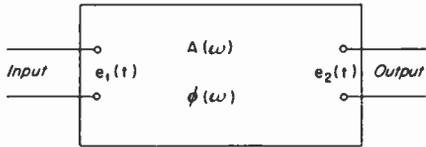


FIG. 15. Transmission medium with input and output quantities expressed as a function of time. The transmission characteristic T of the medium is defined as $T = \frac{e_2(t)}{e_1(t)} = A(\omega) / -\phi(\omega)$, where $A(\omega)$ is the amplitude-frequency relation and $\phi(\omega)$ is the phase-frequency relation.

The behavior of the transmission system or medium may be determined by specifying the input function, expressing the output function analytically, and then comparing the output function of the system with the input function. In making such a comparison it will be convenient to make use of the amplitude-frequency characteristic of the transmission system, $A(\omega)$, and also the phase-frequency characteristic, $\phi(\omega)$, which correspond to the spectra method of representing intelligence. The comparison of the input and output functions indeed yields the amplitude and phase characteristics of the system as a function of frequency.

Let the signal fed into the input terminals of a linear transmission system of Fig. 15 be given by

$$e_1(t) = \sum_{n=-\infty}^{+\infty} |E_{1n}| e^{i\phi_{1n}} e^{in\omega t}$$

for a complex wave form expressible by a Fourier series in which n indicates the order of the harmonic and the subscript 1 denotes the input end of the system. In passing through a linear transmission system, this input function will give rise, at the output terminals, to an output function having the same harmonic content as that of the input. Accordingly, the output function derived from the transmission system

may be expressed as

$$e_2(t) = \sum_{n=-\infty}^{+\infty} |E_{2n}| e^{i\phi_{2n}} e^{jn\omega t}$$

where, in both of the above equations, $|E_n|$ represents the absolute value of the amplitude of the n th harmonic, $e^{i\phi_n}$ represents the phase displacement of the n th harmonic, and $e^{jn\omega t}$ represents the simple harmonic time variation for the n th harmonic component.

The response characteristic of the transmission system (also called the transmission characteristic) is defined as the ratio of the output function to the input function and is given analytically by the relation

$$T = \frac{e_2(t)}{e_1(t)} = \frac{\sum_{n=-\infty}^{+\infty} |E_{2n}| e^{i\phi_{2n}} e^{jn\omega t}}{\sum_{n=-\infty}^{+\infty} |E_{1n}| e^{i\phi_{1n}} e^{jn\omega t}} = \sum_{n=-\infty}^{+\infty} \frac{|E_{2n}|}{|E_{1n}|} e^{i(\phi_{2n} - \phi_{1n})}$$

The transmission characteristic may be resolved into its amplitude characteristic, defined as

$$A(\omega_n) = \frac{|E_{2n}|}{|E_{1n}|}$$

and the phase characteristic

$$\phi(\omega_n) = \phi_{1n} - \phi_{2n}$$

so that, more compactly,

$$T = A(\omega_n) e^{-i\phi(\omega_n)} = A(\omega_n) / -\phi(\omega_n)$$

In specifying the response characteristics of the transmission system, the subscripts n were retained since this designates the frequency of the n th harmonic of the Fourier series of discrete frequency components. In general, however, the response characteristic of a transmission system is a continuous function of frequency (at least between specified limits) and may be expressed by the relation

$$T = A(\omega) e^{-i\phi(\omega)} = A(\omega) / -\phi(\omega)$$

indicating that both the amplitude and the phase are continuous functions of ω within the useful frequency range of the system. The amplitude characteristic of the transmission system is then defined by the relation

$$A(\omega) = \frac{E_2(\omega)}{E_1(\omega)}$$

and the phase characteristic by the relation

$$\phi(\omega) = \phi_1(\omega) - \phi_2(\omega)$$

The amplitude and phase characteristics of the transmission system are intrinsic properties of the system and may be represented graphically (possibly, analytically) as a function of frequency. It is possible to determine the characteristics of the signal obtained at the output terminals, when the character of the input signal is known in addition to the transmission characteristics, $A(\omega)$ and $\phi(\omega)$. Thus, for any

periodic input function, $e_1(t) = \sum_{n=-\infty}^{+\infty} |E_{1n}| e^{i\phi_{1n}} e^{jn\omega t}$, the output function or signal is

specified by

$$\begin{aligned}
 e_2(t) &= e_1(t) \frac{e_2(t)}{e_1(t)} = e_1(t) A(\omega) e^{-j\phi(\omega)} = \sum_{n=-\infty}^{+\infty} |E_{1_n}| e^{j\phi_{1_n}} e^{jn\omega t} A(\omega) e^{-j\phi(\omega)} \\
 &= \sum_{n=-\infty}^{+\infty} |E_{1_n}| A(\omega) e^{j[n\omega t - \phi(\omega) + \phi_{1_n}]} \\
 &= \sum_{n=-\infty}^{+\infty} |E_{1_n}| A(\omega) \cos [n\omega t - \phi(\omega) + \phi_{1_n}]
 \end{aligned}$$

This result shows that, as compared with the input function, the amplitude of the output signal is multiplied by the amplitude characteristic, $A(\omega)$, and the phase of the output signal is delayed by an amount $\phi(\omega)$ compared to that of the input signal.

27. Conditions for Distortionless Transmission. For distortionless transmission, the wave forms of the input and output functions are to be identical, except possibly for a constant multiplying factor and also for a finite delay in the time of transmission, this delay being designated as t_d . The amplitude multiplying constant is designated as K and may have any real positive value.

If the input signal to the transmission system is given by

$$e_1(t) = \text{Re} \sum_{n=-\infty}^{\infty} |E_{1_n}| e^{j\phi_{1_n}} e^{jn\omega t} = \sum_{n=-\infty}^{\infty} |E_{1_n}| \cos (n\omega t + \phi_{1_n})$$

the output signal, expressed in terms of the input signal and the characteristics of the transmission system, is given by

$$e_2(t) = \text{Re} \sum_{n=-\infty}^{\infty} |E_{1_n}| A(\omega) e^{j[n\omega t - \phi(\omega) + \phi_{1_n}]} = \sum_{n=-\infty}^{\infty} |E_{1_n}| A(\omega) \cos [n\omega t - \phi(\omega) + \phi_{1_n}]$$

where the signal is modified by the amplitude factor, $A(\omega)$, and delayed by a lag of $\phi(\omega)$ radians in passing through the transmission system.

If we adopt as the criterion for distortionless transmission the condition that

$$e_2(t) = K e_1(t - t_d)$$

where K is a real, positive constant or scale factor and t_d is the time of delay in the transmission system, the requirement for distortionless transmission is that

$$\sum_{n=-\infty}^{\infty} |E_{1_n}| A(\omega) \cos [n\omega t - \phi(\omega) + \phi_{1_n}] = \sum_{n=-\infty}^{\infty} K |E_{1_n}| \cos [n\omega(t - t_d) + \phi_{1_n}]$$

This condition requires that the relations

$$A(\omega) = K \quad \text{and} \quad \cos (n\omega t) = \cos [n\omega(t - t_d)]$$

be simultaneously satisfied. The first of these two conditions requires that the amplitude-frequency spectrum for the transmission system be constant and independent of frequency. For a constant delay in transmission time, t_d , the second condition requires that the phase-frequency spectrum have a phase displacement for the transmission system that is proportional to frequency. Hence, the conditions to be fulfilled by a transmission system in order that the signal passing through it shall remain free from distortion require that

$$\begin{aligned}
 A(\omega) &= \text{constant} \\
 \phi(\omega) &\propto \omega
 \end{aligned}$$

If, as shown graphically in Fig. 16, these two conditions are fulfilled by the transmission characteristics of the medium through which signals pass, the system will be free from distortion; the wave form will remain unchanged (except possibly for a scale factor) and all frequency components will be effectively delayed by the transmission delay time, t_d .

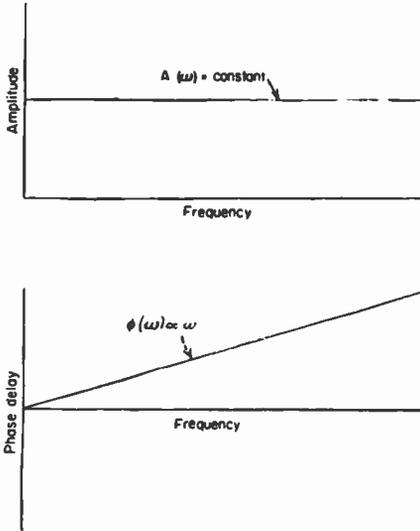


Fig. 16. Amplitude-frequency and phase-frequency characteristics (top and bottom, respectively) for transmission medium which is free from distortion but which effects a time delay for the signal transmitted through it.

mentals of a radio communication system is to apply physiological (and certain aspects of psychological) knowledge for the analytical formulation of intelligence to an electrical system capable of conveying intelligence between two points by means of electromagnetic radiation. The technical applications of radio engineering are based on the fundamentals of electromagnetism, which are briefly summarized in the remainder of this section.

28. Electricity and Magnetism Related to the Structure of Matter. Many experimental observations, particularly within the last century, have conclusively demonstrated the atomic nature of electrical and magnetic phenomena, and the close dependence of electrical effects upon the ultimate structure of matter.

As conceived by the physicist, the atom model is considered to be composed of a stable central core or nucleus spinning on its axis, this compact mass being made up of an aggregation of relatively dense particles called "neutrons" and "protons." Around this central nucleus a definite number of electrons (the number being different for each chemical element) are assumed to revolve in elliptical shells or orbits at distances that are large compared to the size of the nucleus. The revolving electrons are also pictured as spinning on their own axes.

Studies of the structure of matter indicate that all material substances are composed of various aggregations of the fundamental physical particles whose properties, so far as they are known, are briefly summarized in Table 4. Within the past few decades, such great strides have been made in understanding the structure of molecules and atoms that textbooks on the subject can hardly hope to recount current ideas.

References

Principles of Communication:

- Hartley, R. V. L.: *Transmission of Information*, *Bell System Tech. J.*, 7 (3), 535, July, 1928.
 Mertz, P., and F. Gray, *A Theory of Scanning and Its Relation to the Characteristics of the Transmitted Signal in Telephotography and Television*, *Bell System Tech. J.*, 13 (3), 464, July, 1934.
 Osborne, H. S.: *Principles of Electric Circuits Applied to Communication*, *Bell System Tech. J.*, 8 (1), 3, January, 1929.

Mathematical Principles:

- Byerly, W. E.: *Fourier Series and Spherical Harmonics*, Ginn, 1893.
 Carson, J. R.: *Electric Circuit Theory and the Operational Calculus*, McGraw-Hill, 1926.
 Franklin, P.: *Differential Equations for Electrical Engineers*, Wiley, 1934.
 v. Kármán, T., and M. A. Biot: *Mathematical Methods in Engineering*, McGraw-Hill, 1940.

FUNDAMENTALS OF ELECTROMAGNETISM

The first step in the development of an electrical communication system is concerned with the important human physiological characteristics of speech, hearing, and vision, and with analytical methods of expressing intelligence. The second stage in building up the funda-

Fortunately for the electrical engineer the advances in atomic physics have not caused him any concern as to his understanding of electrical phenomena. Nor are the advances in atomic physics likely to produce any radical changes in the theories of electrical sciences; the electrical technologist gets along quite well with the physical models constructed by Faraday and the system of equations crystallized by Maxwell to express known experimental facts, both of which were proposed long before the electron was known. So far as electrical effects are related to the structure of matter, the electrical engineer is concerned primarily only with the outermost shells of electrons in atoms. He may also be concerned with ions—atoms from which electrons have been removed or to which they have been added. The normal atom, with electrons traversing their regular paths, shows no electrical effects. Outside the normal atom, therefore, there exist no electrical effects that are attributable to the position and movements of the interatomic electric charges, although magnetic effects (which are due to the relative spin or rotation of the atomic charges) may exist.

Table 4. Properties of Atomic Particles

Name of particle	Composition	Relative mass	Relative charge	Principal characteristics
Electron.....	Elementary	1	-1	Makes up outer structure of atoms; accounts for the chemical properties of atoms
Positron.....	Elementary	1	+1	Secondary constituent of cosmic rays; too transitory to have found much use
Proton.....	Elementary	1,849 (approx)	+1	Nucleus of H atom and constituent of all atomic nuclei; accounts for charge of nuclei and usually about half of nuclear mass; same as ionized H atom
Neutron.....	Electron-proton	1,849 (approx)	0	Constituent of all nuclei and accounts for appreciable part (usually about one-half) of mass of nuclei
Neutrino.....	?	1	0	Possibly created along with electron and positron when proton changes to neutron or vice versa
Meson.....	Possibly 2 particles with opposite charge similar to electron-proton pair	202	0	Possibly created during nuclear radiation of energy; present in cosmic rays
Deuteron.....	Neutron-proton	3,698 (approx)	1	Ionized heavy H atom
Alpha particle...	2 neutrons and 2 protons	7,396 (approx)	2	Nuclei of He, or He atoms that have lost 2 planetary electrons

The modern view of electricity regards an electric current (at least the conduction and convection components of the total current) as the flow of negative charges in one direction through a given area, plus the flow of positive charges (when these are free to move) through the same area but in the opposite direction. In most cases the negative charges accounting for current flow are due to the motion of electrons, whose properties are given in Table 5. By common agreement, the positive sense of the

Table 5. Properties of the Electron

Rest mass of electron, m_e	9.1066×10^{-31} kg
Charge of electron, e	1.60203×10^{-19} coulomb
	1.60203×10^{-20} abcoulomb
	4.803×10^{-11} statcoulomb
Ratio of charge to mass, (e/m_e)	1.7592×10^{11} coulomb/kg
Size of electron (roughly).....	1.9×10^{-16} m

flow of current is taken as that in which positive charges tend to move or (what is the same thing) opposite to that in which the negative charges move or tend to move. This convention was adopted before the true nature of the electric current was known

and, since (in most cases) the electric current is constituted by the flow of electrons, the anomalous and awkward situation has arisen that the direction of the flow of an electric current is generally in a direction opposite to the flow of charged particles that produce the current! Thus, according to an accepted standard:¹ "The direction of a current through a surface is, by agreement, taken as the direction of the movement of the positive electricity when it is the predominating component in motion and as the direction opposite to the direction of movement of the negative electricity when the latter is the predominating component in motion." For engineering design problems, it makes little difference what the true direction of flow of the charged particles really is, but when this is important, as in the case of rectification, care should be taken to specify whether the direction of flow is that of the electrons or the direction of the flow of positive charges.

The motion of electric charges is produced by the application of electrical forces to material substances. A substance (solid, liquid, or gas) such as any of the metals, carbon, alkaline or saline solutions, or ionized gases, in which a relatively large drift of electrons or other charged particles is produced by a relatively small externally applied electrical force, or in which the charges may travel freely in distributing themselves within the substance, is called a "conductor." A substance (solid, liquid, or gas) such as sulfur, glass, porcelain, mica, rubber, quartz, very pure water, and un-ionized gases, in which measurable amounts of current may be produced (without disrupting the material itself) only by the application of a comparatively large electrical force or in which charges are bound within certain small regions and are not free to distribute themselves readily within the substance, is called a "dielectric" or an "insulator." There is no sharp dividing line between conductors and insulators. The classification of a material into conductors or insulators depends, in some measure, on its application as well as on its inherent physical properties. A substance that might be regarded as a good insulator for some applications might also be regarded as a fairly good conductor for other applications of electrical technology.

The earliest experiments with magnetism were made with lodestones and permanent magnets. There was no definitely known connection between electricity and magnetism until 1820 when Oersted showed that the flow of a steady electric current through a conductor was capable of producing a constant magnetic field about it. In 1833, Faraday showed that the relative motion between a conductor and a magnetic field was able to produce a varying flow of electricity in the conductor. Hence, the experiments of Oersted and Faraday brought together electricity and magnetism, and the work of these two men, supported by the brilliant theoretical and experimental work of Ampère and the outstanding mathematical generalizations of Maxwell, provides the basis of present-day ideas of electromagnetism. Maxwell's equations, forming the foundations of electromagnetic phenomena, are as useful today—and far more important—than when they were formulated about 1865, even though our knowledge of the ultimate causes of electric and magnetic phenomena has undergone radical changes in the past eight decades.

References

- Eddington, A.: "Nature of the Physical World," Cambridge, 1928.
 Haas, A.: "Introduction to Theoretical Physics," London, 1924.
 Houston, W. V.: "Principles of Mathematical Physics," McGraw-Hill, 1948.
 Lorentz, H. A.: "Theory of Electrons," New York, 1909.
 Page, L.: "Introduction to Theoretical Physics," Van Nostrand, 1929.
 Slater, J. C., and N. H. Frank: "Introduction to Theoretical Physics," McGraw-Hill, 1933.
 Stranathan, J. D.: "The 'Particles' of Modern Physics," Blakston, 1942.
 Van Vleck, J. H.: "Theory of Electric and Magnetic Susceptibilities," Oxford, 1932.
 Wilson, W.: "Electromagnetism and Optics," London, 1933.

ELECTROSTATICS

29. Definition. Electrostatics is that branch of science which deals with the laws of electricity at rest. Historically, it is the oldest branch of electricity, but it has

¹ "American Standard Definitions of Electrical Terms," AIEE, 1942.

little direct application to electrical communication, which requires that some electrical quantity vary with time if intelligence is to be transmitted. The ideas originally developed in the study of electrostatics are but special cases of more general relations of electromagnetism.

30. Electric Charges. For all engineering purposes, electrified particles are completely specified by the magnitude and sign of their electric charge and by their rest mass.

An electron is the natural elementary quantity of negative electricity or the natural elementary negative electric charge. The charge (or quantity of electricity) of an electron, represented by the symbol e , is -1.602×10^{-19} coulomb, and the rest mass of an electron is 9.107×10^{-31} kg. Alternatively, one may define the coulomb, the unit of charge for engineering purposes, as the charge whose magnitude is equal to that of 6.242×10^{18} electrons, and the kilogram as the mass represented by 1.098×10^{31} electron rest masses

A proton is the natural elementary quantity of positive electricity or the natural elementary positive electric charge. The charge of a proton is numerically equal to the charge of an electron but is of opposite sign, or $+1.602 \times 10^{-19}$ coulomb. The rest mass of the proton is 1,849 times that of the electron or 1.684×10^{-27} kg.

The quantity of electricity, electric charge, on (or in) a body is the excess of one kind of charge over that of the other. A plus sign indicates an excess of protons or positive ions or, conversely, a deficiency of electrons or other negative charges. Likewise, a minus sign indicates an excess of negative charges or a deficiency of positive charges.

31. Electric Field. Associated with each electric charge is a sphere of influence or a field of force, which tends to accelerate or impart motion to other charges in the field

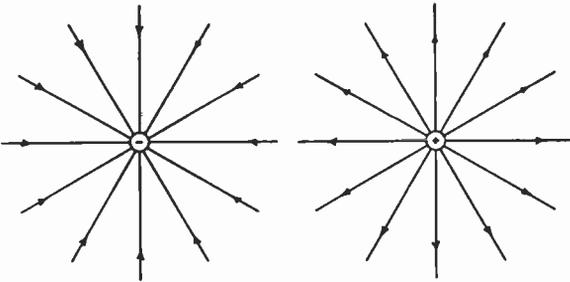


FIG. 17. Representation in a two-dimensional plane of lines of force about isolated electric charges.

of force. Such a field of force is detected only by the force it exerts on charged particles. The charge and its own field of force cannot be separated; indeed, the charge is sometimes looked upon as the point at which the field may be considered to be concentrated.

Since any charge is the result of the excess or deficiency of electrons, the charge, Q , of any substance is given by $Q = \pm ne$ where n is any integer expressing the number of elementary charges involved and e is the numerical value of the electronic charge. The sign of the charge, Q , is positive (+) or negative (-) depending upon whether the charge produces a force on other charges like that of a proton or an electron, respectively.

The sphere of influence, or the field of force, surrounding a charged particle is represented by lines of force that originate at the center of the charge and radiate outward in straight lines in all directions, as shown in Fig. 17. The direction of the lines of force designate the direction of the force that the charge exerts on an arbitrarily selected positive test charge, and the density of the lines of force is proportional to the magnitude of the force.

Since electrical effects in free space are propagated with the velocity of light, the field of any element of charge may be supposed to be emitted continuously from that element and to spread out in all directions from it with the speed of light. Such a view should not be regarded as a "true explanation" of actual physical behavior, of course, but rather as a convenient means of accounting for the fact that electrical effects, due to the motion of charges, at a distance from the charges are observed to take place after a time interval whose magnitude depends upon the distance between the charges and the points of observation. Electrical effects at a distance are retarded in time.

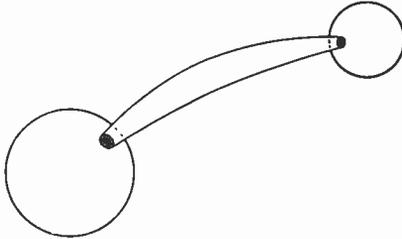


FIG. 18. Tube of force between two charged bodies.

is a curve so drawn that, everywhere, it indicates the direction of the electric intensity, the sense in which the line is described being indicated by an arrowhead on the diagram to show the direction in which the positive test charge would tend to move if placed at the position of the arrowhead.

A bundle of a large, but constant, number of lines of force is known as a "tube" of force. For a single isolated charge, the tube of force is actually a cone, but in most cases it is truly represented as a tube, as shown in Fig. 18.

32. Lines of Force and Tubes of Force. The properties of charged particles may be represented by means of lines of force, a concept to which Faraday attached physical significance. A line of force may be considered as a path indicating the direction of the force applied upon a positive test body (charge) in an electric field. A line of force in an electric field

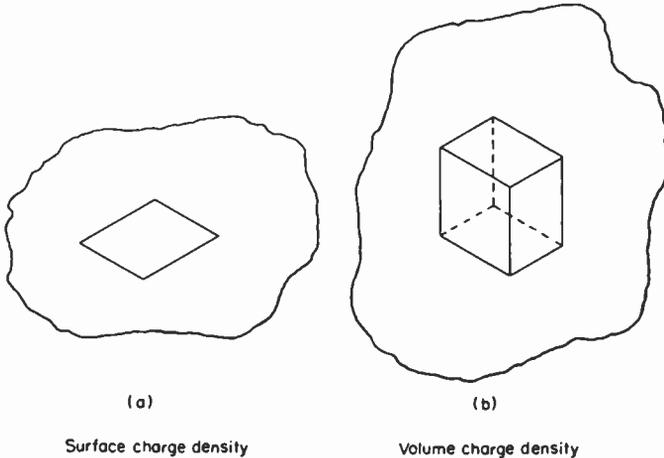


FIG. 19. Surface and volume charge density for electric charges distributed over a surface (left) or throughout a volume (right). The surface and volume are infinitesimal in size.

33. Charge Density. A point charge, *i.e.*, an electrical charge so small that it may be considered as concentrated at a point in space, is the basis of electrostatics, and it is conceivable that electrical problems could be solved in terms of point charges only. When many charges are to be dealt with, however, such a procedure becomes tremendously tedious and cumbersome and, in such cases, the concept of charge density is particularly useful in simplifying the solution of electrical problems. Two kinds of charge density of practical importance may be distinguished.

If the point charges are distributed over a surface area, as shown in Fig. 19a, the surface charge density, σ , is defined as the limiting ratio of the charge to the surface area, as the area approaches zero. If, in mks units,¹ Q is the charge in coulombs and s is the surface area in square meters, the surface charge density in coulombs per square meter is defined to be

$$\sigma = \lim_{\Delta s \rightarrow 0} \frac{\Delta Q}{\Delta s}$$

If the charges are distributed throughout a volume, V , as shown in Fig. 19b, the volume charge density, ρ , is defined as the limiting ratio of the charge to the volume, as the volume approaches zero. If Q is the charge in coulombs and V is the volume in cubic meters, the volume charge density is defined mathematically as

$$\rho = \lim_{\Delta V \rightarrow 0} \frac{\Delta Q}{\Delta V}$$

The charge density is defined at a point on a surface or in a volume, and its magnitude may vary from point to point. The sign of the charge density is that of the charge.

34. Electric Field Intensity. The electric field intensity (also called "electric intensity," "electric field strength," or "electric force") at a point is a vector whose direction is that of the force that would be exerted on a charged particle placed at the point and whose magnitude is equal to the ratio of the force to the charge of the electrified (charged) particle, provided the latter is sufficiently small so that it does not measurably affect or alter the electric field. A vanishingly small charge must, therefore, be employed as the test body to measure the electric field intensity. Hence, the electric field intensity may be defined mathematically as

$$\varepsilon = \lim_{\Delta Q \rightarrow 0} \frac{\Delta f}{\Delta Q}$$

where, in mks units, f is the force in newtons, Q is the charge in coulombs, and ε is the electric field intensity in volts per meter.

35. Coulomb's Law. As early as 1773, Sir Charles Cavendish made the first quantitative measurements in the field of electricity and deduced the relation between the forces exerted by charges, but his results were unpublished and remained unknown for almost a century. In the meantime, in 1785, Charles Augustin Coulomb deduced the same relations between the forces on charges by another approach, and this important relation is now known as Coulomb's law.

Coulomb's law may be stated in words as follows: In a homogeneous, isotropic medium the force of attraction or repulsion between two charges of electricity concentrated at two points that are small compared to the distance between them is proportional to the product of their charge magnitudes, inversely proportional to the square of the distance between the charges, and inversely proportional to a characteristic constant of the medium in which the charges reside, called its "permittivity." The force between unlike charges is one of attraction; the force between like charges is one of repulsion.

Mathematically, Coulomb's law may be stated in the form

$$f = K_1 K_2 \frac{Q_1 Q_2}{\varepsilon l^2} = K_1 K_2 \frac{Q_1 Q_2}{\epsilon_0 \epsilon_r l^2}$$

¹ In this section, electrical quantities are expressed in meter-kilogram-second (mks) units since this system was recommended for adoption after January, 1940, by the International Electrotechnical Commission. Conversion from the mks to the cgs units may be made by means of the conversion factors given in Table 13. The question of rationalization was left open by the IEC so that, at the present time, there is no unanimity as to which set of equations should be free from the unavoidable constant, 4π , representing the number of solid radians in a closed surface. In this section, K coefficients, with suitable primes, are used to replace coefficients involving the factor 4π . By inserting the appropriate value of K , as given in Table 6, the equations in this section are directly applicable to either the unrationalized or the rationalized system.

where, in mks units, f is the force in newtons exerted by the two charges on one another, Q_1 and Q_2 are the two charges in coulombs, $\epsilon = \epsilon_0 \epsilon_r$ is a physical property of the homogeneous isotropic medium known as the permittivity and is measured in farads per meter, ϵ_0 is the permittivity of free space (vacuum) measured in farads per meter, ϵ_r is a numeric expressing the ratio of the permittivity of any substance to the permittivity of free space, l is the distance between the point charges in meters, K_r is a constant which depends upon the units in which the force and charge are measured, and K' , whose value is given in Table 6, is a constant whose value depends upon whether the unrationalized or the rationalized system of equations is employed.

Table 6. Factors for Rationalized and Unrationalized Systems of Units

Factor	Value in unrationalized system	Value in rationalized system
K'	1	$\frac{1}{4\pi}$
K''	$\frac{1}{4\pi}$	1
K'''	4π	1
K''''	1	4π

applies to electric currents (which are charges in motion) so long as the medium is homogeneous and isotropic.

36. Conservative Properties of Static Fields. Problems in electrostatics may be solved by the fundamental relations expressed by Coulomb's law, which deals with forces that have direction as well as magnitude. The calculation of directed forces is a laborious and tedious undertaking, so that, except in the simplest configurations of charges, calculations by Coulomb's law become too cumbersome for general use. The concept of potential greatly simplifies problems in electricity by eliminating the need to consider direction (except in a positive and negative sense). Since the potential concept is applicable only to conservative fields, it is necessary to ascertain that the electrostatic field is conservative, *i.e.*, that the energy of an electrostatic field is a function of position only and does not depend upon the path that the charge takes between the two points.

According to Coulomb's law and the definition for work, the increment of work done in moving a charge a distance dr is given by the equation¹

$$dw = -f \cos \theta \, dl = -f \cdot dr = -(K_r K' Q_1 Q_2 / \epsilon r^2) \, dr$$

¹ When vector quantities are encountered, it is frequently necessary to express physical quantities in terms of the component of the vector that is along a specified path or is normal to a specified surface. For example, the increment of work is equal to the product of the distance through which a body moves by the component of force in the direction of motion, or, $dw = f \cos \theta \, dl$, where f is the force, dl is the element of length in the direction of motion, and θ is the angle between the direction of the force and the direction of motion. Thus, $f \cos \theta$ is that component of the force in the direction of motion along dl . In cases such as this, the desired vector (in this example, the element of work) may be expressed in any of the alternative ways

$$f \cos \theta \, dl = f \cos (\mathbf{f}, \mathbf{l}) \, dl = \mathbf{f} \cdot \mathbf{dl}$$

where the notation (\mathbf{f}, \mathbf{l}) represents the angle between the direction of \mathbf{f} and that of \mathbf{l} . The dot (\cdot) is used to represent the cosine of this angle, or $\cos \theta$.

The vector \mathbf{f} may be resolved into its components along any set of three coordinates in space. Each such component is, of course, a vector quantity and may be specified in terms of its component vectors in the direction of each coordinate. Thus, the vector \mathbf{f} may be resolved, for example, into its rectangular coordinate components, f_x , f_y , and f_z , as in the expression

$$\mathbf{f} = f_x + f_y + f_z$$

A vector may also be specified in terms of a unit vector in the direction of each coordinate and a scalar

the negative sign being used because positive work is done in moving the charge in the $-r$ direction. The total amount of work done in moving the charge from any point A to any other point B whose distances from any reference point are r_A and r_B , respectively, is then the integral of dw between the two points or

$$w = w_A - w_B = \int_B^A dw = - \int_{r_B}^{r_A} \left(\frac{K_c K' Q_1 Q_2}{\epsilon r^2} \right) dr = - \left(\frac{K_c K' Q_1 Q_2}{\epsilon r} \right) \left(\frac{1}{r_A} - \frac{1}{r_B} \right)$$

This result is independent of the path taken between A and B and depends only upon the positions of the end points of the path traversed. Therefore, the electrostatic field is conservative.

37. Electrostatic Potential. The electric potential difference between two points in an electrostatic field is equal to the work associated with the transfer of unit quantity of positive charge from one point to another. The maximum amount of work is done by (or on) charges that move along the lines of electric force, whereas no work is done by (or on) a charge that moves perpendicular (normal) to lines of electric force. If A and B are the two points between which a unit positive charge is transferred, then B is at a higher potential than A if external energy is required to transfer unit positive charge from A to B .

The great advantage of the concept of potential is that, through its use, vector calculations of the forces acting on charges may be resolved into scalar calculations which are much simpler to manipulate. From the definition given above, it is evident that there exists only a means for determining the difference or change of potential between two points and not for determining the absolute value of the potential of any point. If it is possible to establish as an arbitrary reference an equipotential surface, i.e., a surface normal to the lines of force upon which no work is required to transfer a positive charge from one point to another on the surface, this surface may be established or defined as being at zero potential. The potential of any point may then be specified relative to the equipotential surface arbitrarily designated as being at zero potential. In engineering work the earth is often arbitrarily taken as the zero reference equipotential surface, and potential differences are then measured relative to the potential of the earth's surface. In other cases, particularly in theoretical work, it is convenient to assign zero potential to a sphere of infinite radius, since the force produced at a point infinitely removed from a charge is zero.

The increment of work per unit charge that is done in moving the charge a distance dl in an electric field is $dw_Q = - (f/Q) \cos \theta dl$. But by definition, $\epsilon = f/Q$ and, therefore, $dw_Q = -\epsilon \cos \theta dl$. The difference in potential between two points A and B is the integral along the path from A to B , or

$$E_{AB} = V_A - V_B = - \int_B^A \epsilon \cos \theta dl = - \int_B^A \epsilon \cdot dl$$

quantity representing the magnitude of the vector component along each coordinate. Thus, if i , j , and k are unit vectors in the x , y , and z directions, respectively, and if f_x , f_y , and f_z are the magnitudes (scalar quantities) of f in these three directions, the vector may be expressed in the alternative form

$$f = if_x + jf_y + kf_z$$

The vector may represent a quantity flowing through a surface. For example, the amount of liquid that passes outward through the element of surface, ds , in unit time at any point on the surface, is the product of the outward normal (or perpendicular) component of the vector F and the surface element, ds . If n is a unit vector normal to the surface element, ds , the outward flow of liquid may be expressed in any of the alternative ways

$$F \cos \theta ds = F \cos (F, n) ds = F \cdot n ds = F_n ds$$

where (F, n) represents the angle between the direction of F and the surface normal, n . The dot (\cdot) expresses the cosine of the angle between F and n , and F_n is the component of F perpendicular to the surface element.

In some texts, especially those stressing vector notation, vector quantities are printed in boldface (heavy) type, as has been done for illustration in this footnote. In other cases, as in the main text of this section, ordinary light-face characters are employed. In this section script letters have been used for those electrical quantities (as \mathcal{E} , \mathcal{D} , \mathcal{H} and \mathcal{B}) which are regarded as vectors in space.

where V_A is the absolute potential of point A , V_B is the absolute value of potential of point B , $E_{AB} = V_A - V_B$ is the difference in potential between points A and B , ε is the electric intensity in the region along the path from A to B , and θ is the angle between the direction of ε and that of dl . Although numerical values for V_A and V_B are not known, the difference $V_A - V_B$ can be ascertained. In the mks system of units, unit potential difference is the joule per coulomb, which has been given the name "volt."

If the electrostatic field is made up of concentrated charges, $Q_1, Q_2, Q_3, \dots, Q_n$ in a homogeneous isotropic medium whose permittivity is ϵ , and in which the distances of these charges from a point P are $r_1, r_2, r_3, \dots, r_n$ and if the potential of a sphere infinitely distant is regarded as the reference or zero potential, then, with respect to the zero reference potential, the potential of the point P is

$$\begin{aligned} E_P &= - \int_{\infty}^{r_1} \frac{K_e K' Q_1}{\epsilon r_1^2} dr - \int_{\infty}^{r_2} \frac{K_e K' Q_2}{\epsilon r_2^2} dr - \int_{\infty}^{r_3} \frac{K_e K' Q_3}{\epsilon r_3^2} dr - \dots - \int_{\infty}^{r_n} \frac{K_e K' Q_n}{\epsilon r_n^2} dr \\ &= - \frac{K_e K'}{\epsilon} \left(\frac{Q_1}{r_1} + \frac{Q_2}{r_2} + \frac{Q_3}{r_3} + \dots + \frac{Q_n}{r_n} \right) \\ &= - \frac{K_e K'}{\epsilon} \sum_{n=1}^m \frac{Q_n}{r_n} \end{aligned}$$

In practice such an expression as that given above is applicable only if a very small number of charges is encountered, for the calculations become tedious and cumbersome if many charges must be taken into account.

When many charges make up an electrostatic field, it is often convenient to disregard the individual charges and deal with their net or over-all effects as expressed by the density of charge. If σ is the surface charge density and ρ is the volume charge density, then the charge distributed over an elementary area, ds , is $dQ = \sigma_n ds$ and the charge distributed throughout an elementary volume, dv , is $dQ = \rho dv$. In the case of a continuous distribution of charge throughout a volume and also over a surface, the potential of the point P (with respect to zero potential at a point infinitely remote) whose distance from the volume and the surface is r , is given by

$$E_P = \int \int_s \frac{dQ}{r} + \int \int \int_V \frac{dQ}{r} = \int \int_s \frac{\sigma_n ds}{r} + \int \int \int_V \frac{\rho dv}{r}$$

The electrostatic potential (defined as the work per unit charge done in, or available for, transporting electric charges) is applicable to static fields. The concept of potential may be generalized and applied to fields in which the charges vary with time. The potential field is a scalar field whereas the field of force is a vector field.

38. Potential Gradient. The gradient of a scalar field at a point is a vector whose magnitude is equal to the space rate of change of the scalar field in the direction of greatest increase and whose direction is that of the greatest increase. This general definition may be applied to the potential field as follows:

From the relations given in Sec. 37 it is easy to show that $dE = -\varepsilon \cos \theta dl$ from which $\varepsilon \cos \theta = -(dE/dl)$. If the direction of the field of force exerted on a small positive test charge is in the direction of the element of path length, dl , then $\cos \theta = 1$. For this condition the change of potential is a maximum with respect to a change of position. The maximum rate of change of potential with position is a vector quantity whose direction is at right angles to the equipotential surface or along the electric lines of force; it is called the "potential gradient" and is abbreviated "grad E ." Hence, from the results above,

$$\text{grad } E = \frac{dE}{dl} = -\varepsilon$$

In the mks system of units, the gradient is measured in volts per meter.

The potential gradient may be resolved into orthogonal components. In rectangular coordinates,

$$\text{grad } E = \nabla E = \left(i \frac{dE_x}{dx} + j \frac{dE_y}{dy} + k \frac{dE_z}{dz} \right) = -(i\epsilon_x + j\epsilon_y + k\epsilon_z)$$

where *i*, *j*, and *k* are unit vectors in the *x*, *y*, and *z* directions, respectively, and the components of *E* and ϵ in each of the rectangular coordinate directions are indicated by the subscript *x*, *y*, or *z*.

A pictorial representation of potential gradient resolved into its orthogonal components in rectangular coordinates is given in Fig. 20. The potential gradient is, strictly speaking, applicable to the field intensity only in the case of an electrostatic field.

39. Electric Flux. Consider a surface element at a point in an electric field for which the electric intensity is ϵ . If θ represents the angle between the direction of ϵ and the normal to the surface, *n*, then, as shown in Fig. 21 the component of ϵ perpendicular to the surface is $\epsilon \cos \theta = \epsilon \cos (\epsilon, n) = \epsilon \cdot nds = \epsilon_n ds$. The product of the normal component of the electric field intensity and the surface is called the "electric flux."

The electric flux, represented by the symbol ψ , may be represented by lines that are drawn in the same direction as the lines of electric intensity but of such spacing that their density (in lines per unit area) perpendicular to the direction in which the lines run is equal to the dielectric flux density at the point at which the flux is measured. From a physical point of view, the electric flux may be regarded as the number of lines of electric field that cross perpendicular to a surface; it is the total number of such lines and not the density in lines per unit area.

Electric flux is a scalar quantity, the sign of which depends upon the assumption made concerning the positive direction of the normal. With a closed surface, the outward normal is taken as positive. For an open surface, if the positive sense is that in which the periphery of the surface is described by the rotation of a right-hand screw normal to the surface, the positive direction of the normal is that in which the screw advances.

In the mks system of units, the electric flux, ψ , is measured in coulombs, the same unit in which charge is measured.

In the unrationalized system, the magnitude of the units of electric flux and charge are the same, but in the rationalized system the unit of electric flux is $1/4\pi$ as large as in the unrationalized system.

40. Electric Displacement; Electric Flux Density. The electric displacement, or the electric flux density, at any point in an isotropic dielectric medium is a vector having the same direction as that of the electric field intensity, and a magnitude that is ϵ times that of the electric field intensity. If, in mks units, ϵ is the electric field

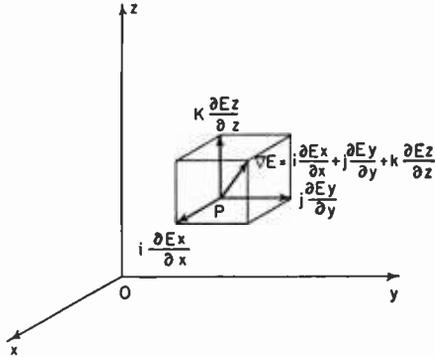


FIG. 20. Potential gradient at the point *P* in terms of its orthogonal components, expressed in rectangular coordinates.

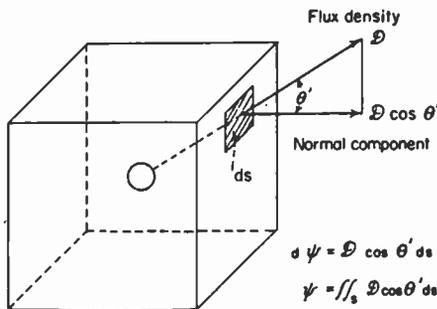


FIG. 21. Determination of electric flux diverging from closed surface. The total flux is the sum of all components normal to a closed surface.

intensity in volts per meter, \mathfrak{D} is the electric displacement in coulombs per square meter, and ϵ is the permittivity of the dielectric medium in farads per meter, then the electric flux density in an isotropic dielectric is

$$\mathfrak{D} = K''\epsilon\mathcal{E}$$

Thus, the displacement takes into account the permittivity properties of the medium through which the lines of electric flux flow; the properties of the medium are inherent in the displacement vector itself.

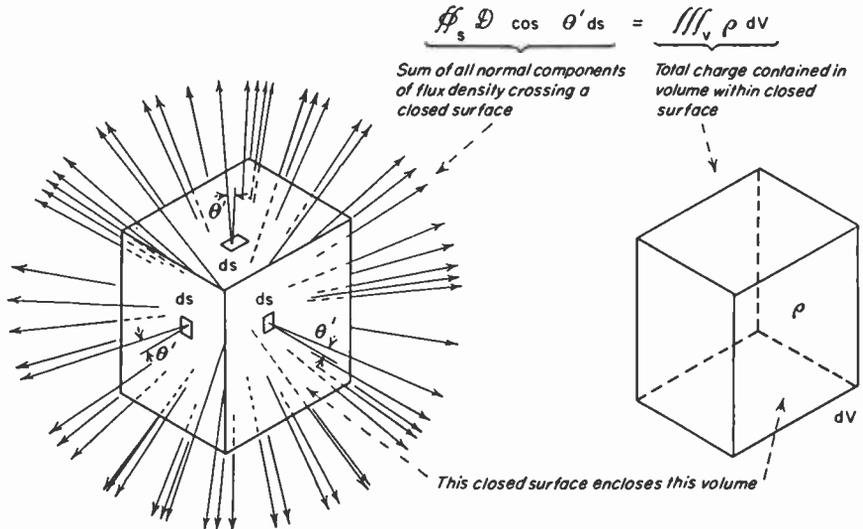


FIG. 22. Total electric flux diverging from a closed surface (left) is equal to the total charge enclosed by that surface.

If \mathfrak{D} represents the electric displacement vector making an angle θ with the surface through which the electric field passes, then the element of electric flux crossing an element of surface is

$$d\psi = \mathfrak{D} \cos \theta ds = \mathfrak{D} \cos (\mathfrak{D}, s) ds = \mathfrak{D}_n ds = \mathfrak{D} \cdot ds$$

The total flux is the integral over the closed surface through which flux passes or

$$\psi = \int \int_s \mathfrak{D} \cos \theta ds = K'''Q$$

where, in mks units, ψ is the electric flux in coulombs, \mathfrak{D} is the displacement in coulombs per square meter, and ds is an element of surface, expressed in square meters. This equation merely states that the total flux diverging from a closed surface is proportional to the total charge enclosed by the surface, as shown in Fig. 22.

41. Dielectric Polarization. For a material dielectric the displacement, \mathfrak{D} , may be split into two parts

$$\mathfrak{D} = \epsilon\mathcal{E} = \epsilon_0\epsilon_r\mathcal{E} = \epsilon_0\mathcal{E} + (\epsilon_r - 1)\epsilon_0\mathcal{E}$$

The first term, $\epsilon_0\mathcal{E}$, is the displacement for free space. The second term, $(\epsilon_r - 1)\epsilon_0\mathcal{E}$, represents an actual displacement or shift of bound electric charges within the material dielectric itself. This latter displacement or shift is customarily measured in units of charge per unit area, or in mks units, in coulombs per square meter; it is called the "polarization" of the dielectric and is designated by the symbol \mathcal{P} . Since the unit of

displacement is K' times the unit of charge, $(\epsilon_r - 1)\epsilon_0 = 4\pi$ so that

$$\mathfrak{D} = \epsilon_0 \mathcal{E} + 4\pi \mathcal{Q}$$

42. Gauss's Theorem. Coulomb's law is one of the most important of all electrical laws and one way of expressing the relation between charge and force has already been given for an isotropic, homogeneous medium. The inverse square law, already given, is the form that is most convenient in providing a quantitative idea of the forces between isolated electric charges, but Coulomb's law may also be given in alternative forms that bring out certain additional general and important properties of charged bodies and facilitate many kinds of calculations. Gauss's theorem is one highly important alternative expression for Coulomb's law. The diagram of Fig. 22 illustrates the theorem.

Gauss's theorem makes use of the concept of electric flux. The theorem may be stated in words as follows: The integral, over any closed surface, of the normal component of the electric flux density or displacement is equal to the charge contained within the surface. In mathematical notation this may be expressed as

$$\psi = \oint_s \epsilon \mathcal{E} \cos \theta \, ds = \oint_s \mathfrak{D} \cos \theta \, ds = \int \int \int_v \rho \, dV = K''Q$$

where \mathfrak{D} is the electric flux density in coulombs per square meter, \mathcal{E} is the electric field intensity in volts per meter, θ is the angle between the direction of \mathfrak{D} or \mathcal{E} and the normal to the surface, ds is an element of surface expressed in square meters, ρ is the volume charge density in coulombs per cubic meter, and dV is an element of volume expressed in cubic meters.

The above equation is the integral form of Gauss's theorem, which may be derived as follows. If the electric displacement, \mathfrak{D} , crosses a surface element, ds , the normal component of flux density crossing the surface is

$$d\psi = \mathfrak{D} \cos \theta \, ds = \epsilon \mathcal{E} \cos \theta \, ds = (Q \cos \theta / r^2) \, ds$$

But $[(\cos \theta) / r^2] \, ds$ is an elementary solid angle, $d\Omega$, and, hence, $d\psi = Q \, d\Omega$. Hence the total flux is the integral over a closed surface, and since there are 4π radians in a closed surface, the flux is given by

$$\psi = \oint_s Q \, d\Omega = K''Q$$

43. Equations of Laplace and Poisson. Gauss's theorem may be put into differential as well as integral form to obtain either the Laplace or the Poisson equation. The differential form may be obtained by resolving the electric flux into three orthogonal components, expanding each component in a Taylor's series expansion and forming the algebraic sum of the linear terms in this expansion for each orthogonal component, over a closed surface. If the volume enclosed is free from charge, the result of such an operation is the scalar quantity given by Laplace's equation,

$$\text{div } \mathfrak{D} = \frac{d\mathfrak{D}_x}{dx} + \frac{d\mathfrak{D}_y}{dy} + \frac{d\mathfrak{D}_z}{dz} = 0$$

whereas if the enclosed volume has a volume charge density, ρ , the equation of Poisson results,

$$\text{div } \mathfrak{D} = \frac{d\mathfrak{D}_x}{dx} + \frac{d\mathfrak{D}_y}{dy} + \frac{d\mathfrak{D}_z}{dz} = K'''\rho$$

The equations of Laplace and Poisson may be expressed in terms of potential as well as of electric field intensity since $\mathcal{E} = -(dE/dl)$. Hence the above equations become

$$\begin{aligned} \frac{d^2 E_x}{dx^2} + \frac{d^2 E_y}{dy^2} + \frac{d^2 E_z}{dz^2} &= 0 \\ \frac{d^2 E_x}{dx^2} + \frac{d^2 E_y}{dy^2} + \frac{d^2 E_z}{dz^2} &= -\frac{K'''\rho}{\epsilon} \end{aligned}$$

Like the integral forms for expressing Gauss's theorem, the equations of Laplace and Poisson are merely alternative ways of expressing the fundamental laws of charge which were first quantitatively observed by Cavendish and Coulomb.

44. Divergence. Divergence is a concept that is useful in describing the flow of a quantity. The divergence of a vector field at any point may be defined as a scalar quantity equal in magnitude to the limiting ratio of the outward flux passing through a closed surface that surrounds the point, to the volume enclosed within the surface, as the volume approaches zero. Physically the operation denoted by divergence yields a measure of the flux produced per unit volume, at a point. When flux flows away from a point, divergence is indicated; when it flows toward a point, negative divergence (sometimes called "convergence") is indicated.

If, in a vector field, the vector, A , is expressed in terms of its three rectangular orthogonal coordinates, A_x , A_y , and A_z , given as functions of x , y , and z coordinates, respectively, the divergence of the vector field is a scalar quantity formed by the sum of the derivatives of each orthogonal component in the direction of its axis, or

$$\begin{aligned} \operatorname{div} A \equiv \nabla \cdot A \equiv (\nabla A) &= \frac{dA_x}{dx} + \frac{dA_y}{dy} + \frac{dA_z}{dz} \\ &= \left(i \frac{d}{dx} + j \frac{d}{dy} + k \frac{d}{dz} \right) (iA_x + jA_y + kA_z) \end{aligned}$$

As an example of this concept, the divergence of the electric intensity, which is proportional to the volume density of charge at the point, may be cited.

45. Curl. Curl is a concept useful in describing the rotation or swirl of a flowing quantity. The curl (also called "rotation," especially by European authors) of a vector field at a point is a vector whose magnitude is equal to the limiting ratio of the circulation around a surface element on which the point is located, to the area of the surface as the latter approaches zero, provided the surface is so oriented as to give the maximum values of circulation. The positive direction of this vector is that traveled by a right-hand screw turning about an axis normal to the surface element when the turning of the screw gives a positive value to the circulation or rotation.

If the vector, A , of a vector field is expressed in terms of its three rectangular components, A_x , A_y , and A_z , given as functions of x , y , and z , respectively, the curl of the vector field is the vector sum of the partial derivatives of each component with respect to the axes perpendicular to it. Mathematically,

$$\operatorname{curl} A \equiv \nabla \times A$$

$$= \begin{vmatrix} i & j & k \\ \frac{\partial}{\partial x} & \frac{\partial}{\partial y} & \frac{\partial}{\partial z} \\ A_x & A_y & A_z \end{vmatrix} = i \left(\frac{\partial A_z}{\partial y} - \frac{\partial A_y}{\partial z} \right) + j \left(\frac{\partial A_x}{\partial z} - \frac{\partial A_z}{\partial x} \right) + k \left(\frac{\partial A_y}{\partial x} - \frac{\partial A_x}{\partial y} \right)$$

where i , j , and k are unit vectors along the x , y , and z axes, respectively.

As an example of the curl concept, the magnetic intensity at a point within an electric conductor, which is proportional to the current density, may be cited.

MAGNETOSTATICS

46. Definition. Magnetostatics is the branch of science that deals with the laws of permanent magnets. Prior to the discovery by Oersted in 1820 of the relation between electricity and magnetism, the study of magnetostatics was pursued independent of that of electrostatics, for in the static case no relation between electrostatics and magnetostatics can be made at all apparent. With the discovery that the flow of electric currents produced magnetic fields, there arose an attempt to explain all magnetic phenomena in terms of the flow of an electric current. Some authors have justified this point of view on the ground that, although the elementary electric charge

has been discovered, no corresponding magnetic charge or pole has been observed as an entity of nature.

There is ample evidence that magnetic effects are the result of the flow of an electric current and possibly some justification for omitting a treatment of magnetostatics. Such a course overlooks the fact that many existing texts already employ the concepts of magnetostatics; it also neglects the historical approach which has given rise to much confusion in the establishment of electrical units.

47. Magnets. A magnet is a body or substance that produces a magnetic field external to itself. The first known magnets were natural magnets of lodestone, but today magnets for practical purposes are made of iron, steel, or alloys in which the principal constituents are usually iron, nickel, cobalt, and perhaps small proportions of other metals such as aluminum or copper.

A magnetic field is a vector field of force imparting, or tending to impart, motion to fine particles of a magnetic substance (*e.g.*, iron filings) situated in the magnetic field.

48. Magnetic Pole. The magnetic poles (or simply the poles) of a magnet are those portions of a magnet toward which the external magnetic field tends to converge or from which it tends to diverge. Magnetic poles always occur in pairs; they cannot be isolated as can electric charges. The size and position of the poles of a magnet depend upon its shape and also upon its magnetic characteristics. For a magnet that is long and slender (as a steel needle), the poles are located at small regions near the ends. In such a case the poles are virtually points, and it is convenient to regard them as concentrated at points.

A study of magnetostatics may be built up on the hypothesis that magnetic poles can be considered to exist in complementary pairs. The mathematics of magnetostatics then becomes very similar to the mathematics for electrostatics, but a different physical interpretation to the equations is required because isolated magnetic poles have not been identified.

49. Coulomb's Law for Magnetic Poles. Experiment shows that for long, thin, magnetized needles, the poles may be regarded as being concentrated at points near their ends. The regions at which the magnetic effects are concentrated may then be regarded as isolated magnetic poles each of which establishes its own spherical magnetic field of force, much as isolated charges establish their spherical electrical field of force. This point of view is justified by the mathematical convenience of the equations of magnetostatics (and the similarity of the equations of magnetostatics with those of electrostatics) rather than by the precision with which this view accords with physical facts. If such a view is adopted, the force between magnetic poles obeys the inverse square law, and Coulomb's law may be applied to magnetic poles as well as to electric charges.

Coulomb's law for magnetic poles may be stated as follows: In a homogeneous, isotropic medium, the force between two magnetic poles that are small compared to the distance between them is proportional to the product of the magnitudes (or strengths) of the two poles, inversely proportional to the square of the distance between them, and inversely proportional to a characteristic property of the medium in which the poles reside, called its "permeability." The force between unlike poles is one of attraction; the force between like poles is one of repulsion.

Mathematically, Coulomb's law for magnetic poles may be stated as

$$f = K_m K' \frac{M_1 M_2}{\mu l^2} = K_m K' \frac{M_1 M_2}{\mu_0 \mu_r l^2}$$

where, in mks units, f is the force in newtons exerted between the two poles, M_1 and M_2 are the magnitudes of the pole strengths of the two magnetic poles, $\mu = \mu_0 \mu_r$ is a physical property of the homogeneous, isotropic medium in which the poles reside, called the "permeability" of the medium, μ_0 is the permeability of free space (vacuum) measured in henrys per meter, μ_r is a numeric expressing the ratio of the permeability of any material substance to that of free space, l is the distance between the two poles

in meters, K_m is a constant relating pole strength and mechanical units of force, and K' is a constant which adapts the equation to the rationalized or unrationalized system and whose value is given in Table 6.

With the exception that no free, isolated, magnetic poles exist (so that, for example, there is no magnetostatic equivalent of the electrostatic charge density), the reasoning developed for electric charges applies to magnetic poles. Thus, Gauss's theorem and the equation of Laplace apply to magnetic poles as well as to electric charges, but Poisson's equation becomes identical with Laplace's since there is no magnetic "pole density."

50. Pole Strength. The strength of a magnetic pole that can be considered as concentrated at a point is measured by the force exerted on the pole when it is placed in a magnetic field of known intensity in a vacuum. A unit magnetic pole is one which is concentrated at a point and which has such strength that, when placed at a unit distance from an exactly similar pole, will experience a unit-repelling force. In the cgs electromagnetic units, the unit of pole strength is the pole, but this unit has no name in the cgs electrostatic or in the mks systems of units.

51. Magnetic Field Intensity. The magnetostatic field intensity due to a magnetic pole is defined as a vector whose magnitude is equal to the force per unit magnetic pole strength exerted on a test body, and whose direction is the same as the direction of the force exerted on the test body. The test body is a concentrated north-seeking pole of such small strength (strictly, infinitesimal strength) that its presence does not alter the magnetic field that it is intended to measure.

The magnetic field intensity may be defined by the magnetostatic relation,

$$\mathcal{H} = \lim_{\Delta M \rightarrow 0} \frac{\Delta f}{\Delta M}$$

where, in the mks system of units, \mathcal{H} is measured in ampere-turns per meter, f is the force in newtons, and M is the pole strength. This equation is analogous to the electric field intensity produced by static electric charges. The magnetic field intensity may also be defined by a dynamic relation in terms of current flow (see Sec. 68).

The direction and magnitude of the magnetic field intensity may be represented by lines of magnetic force, just as lines of electric force represent the direction and magnitude of the electric field intensity.

52. Magnetic Flux. The magnetic flux, ϕ , produced by magnetic poles, through a surface may be regarded as the number of lines of magnetic force that cross perpendicular to the surface. It is a scalar quantity, derived in a way completely similar to that by which electric flux is derived. In the mks system of units, the magnetic flux is measured in webers.

53. Magnetic Induction; Magnetic Flux Density. The magnetic induction, or the magnetic flux density, at any point in a magnetic field, is a vector whose magnitude is μ times that of the magnetic field intensity, \mathcal{H} , and whose direction is that of the magnetic field intensity. If \mathcal{H} is the magnetic field intensity and μ is the permeability of the magnetic medium, then the magnetic induction (or magnetic flux density) is given by the relation

$$\mathcal{B} = \mu\mathcal{H}$$

In mks units the magnetic induction is measured in webers per square meter.

In terms of the magnetic flux density, the magnetic flux is given by

$$\phi = \int_s \mathcal{B} \cos \theta \, ds$$

where θ is the angle between the direction of \mathcal{B} and the normal to the surface element. Over a closed surface,

$$\phi = \oint_s \mathcal{B} \cos \theta \, ds = 0$$

This relation may be regarded as a result of the absence of free, isolated magnetic

poles. Permanent magnets, however small, are always found in equal and opposite pairs. Therefore, for any volume containing a permanent magnet, as many lines of magnetic force leave the surface of enclosure as enter it; so the net or resultant flux is zero. In mks units, ϕ is the magnetic flux in webers, \mathcal{G} is the magnetic flux density in webers per square meter, ds is an elementary surface area expressed in square meters, and θ is the angle between the direction of the flux density and the normal to the element of surface, ds .

In differential form, Gauss's theorem for magnetic flux density is given by

$$\frac{d\mathcal{G}_x}{dx} + \frac{d\mathcal{G}_y}{dy} + \frac{d\mathcal{G}_z}{dz} = 0$$

In words, the divergence of magnetic flux from a closed surface is zero.

54. Magnetization. Magnets are regarded as being produced by the proper alignment or orientation of elementary magnets (of crystalline dimensions) within the crystalline structure of the magnetic material. In the natural state the elementary magnets are at random orientation; hence, the substance shows no magnetic properties. As more and more of the elementary magnets assume the same orientation, the magnetism increases, up to the limit attained when all elementary magnets are aligned in the same direction.

The magnetic properties of material substances may be separated from those of free space in the same way in which dielectric polarization was separated from the permittivity of free space. Thus, we may write

$$\mathcal{G} = \mu\mathcal{H} = \mu_0\mu_r\mathcal{H} = \mu_0\mathcal{H} + (\mu_r - 1)\mu_0\mathcal{H} = \mu_0\mathcal{H} + 4\pi\mathfrak{M}$$

where \mathfrak{M} is the intensity of magnetization which, in the mks system, is measured in webers per square meter.

55. Permeability and Reluctivity. Permeability, designated by the symbol μ , is the property of an isotropic medium which determines, under specified conditions, the magnitude relation between magnetic induction and magnetic intensity (also called "magnetizing force") of the medium.

Under the specified conditions, permeability is measured as the ratio of the magnetic induction to the magnetic intensity. Mathematically, the permeability is defined to be

$$\mu = \frac{\mathcal{G}}{\mathcal{H}}$$

Reluctivity of a medium is the reciprocal of its permeability. Mathematically,

$$\nu = \frac{1}{\mu} = \frac{\mathcal{H}}{\mathcal{G}}$$

ELECTRODYNAMICS

56. Definition. Thus far we have dealt with electric charges and magnetic poles in the static case; *i.e.*, we have dealt with charges and poles free from motion relative to the reference coordinate system. So long as only static conditions were studied, there were no discernible connections between electricity and magnetism. With the discovery, in 1820, by Oersted (quickly repeated by Ampère) that the flow of an electric current produced a magnetic field, and the further discovery by Faraday in 1831 that varying magnetic field was capable of producing an induced electromotive force, the connection between electricity and magnetism was clearly and firmly established, and from then on the unified subject of electromagnetism was studied.

Electromagnetism may be defined as the study of electricity (electric charges) in motion, the magnetic effects produced by the flow of electricity, and, conversely, the electrical effects arising from the relative motion of a magnetic field and path through which electric current may pass or flow. The current may be a conduction current or a displacement current. In the former case, electromagnetic energy may be

guided between two points connected by a conductor. In the latter case, electromagnetic energy may be transferred from one point to another through a dielectric.

The term "electrodynamics" (or electrokinetics) is used in place of electromagnetism, particularly when the motional aspect of the electric charges is emphasized, as in distinguishing electrostatics from electrodynamics. The term "electromagnetism" usually includes electrostatics and, hence, is often regarded as a more encompassing term than "electrodynamics."

57. Electric Current. An electric current through a surface is the time rate at which positive or negative charges pass through it. If both positive and negative charges are simultaneously passed through the surface, the current is the time rate of passage of the algebraic sum of the two. Three types of current flow are recognized: conduction current, convection current, and displacement current.

A conduction current comprises the movement of negative charges (electrons) exclusive of any movement due to the transportation of negative electricity by masses larger than electrons. A convection current is a current in which the charges are electrified particles larger than electrons. The displacement current through any surface in an isotropic dielectric is proportional to the time rate of change of the displacement flux through the surface.

In accordance with these definitions, the conduction and convection currents may be regarded as the flow of physical particles through a surface, the magnitude of the current being proportional to the net flow of the charged particles transported through the surface. No charge can pass through an ideal dielectric, and yet the time-varying charges accumulating on the plates of a capacitor give every indication that their physical presence is equivalent to the flow of a time-varying current through the dielectric. The magnitude of this displacement current is proportional to the time rate of change of the electric flux between the two conductors of the capacitor. But such a time rate of change of electric flux exists in the case of conduction and convection currents as well. Instead of visualizing the electric current as the flow of actual particles through a surface, a more fundamental and basic point of view is to regard the electric current as the time rate of change of the electric flux produced by moving charges. If the charges are able to pass through a surface, then the effect is the usual one associated with a conduction current or with the continuous passage of charged particles. On the other hand, if the charges are prevented from crossing a surface (as in the case of the dielectric in a capacitor), the flow of displacement current is still permissible without any change in fundamental concepts. In any case, the total current flowing at any instant of time is the sum of the conduction, convection, and displacement components of current.

If Q represents the charge of moving charged particles, the conduction (or convection) current is given by

$$I_c = \frac{dQ}{dt}$$

whereas if $\psi = K''Q = Q/K'$ is the electric flux produced by the charge Q and flowing through a surface, the displacement current through this surface due to the variation of the electric flux is

$$I_d = K' \frac{d\psi}{dt}$$

The total current, composed of conduction (or convection) and displacement components, is

$$I_t = I_c + I_d = \frac{dQ}{dt} + K' \frac{d\psi}{dt}$$

In mks units, the charge and flux are measured in coulomb units, and the current (all components, individually or collectively) is measured in coulombs per second or in amperes.

58. Current Density. Current density at a point is a vector whose direction is the same as that of the current flow and whose magnitude is equal to the limiting ratio of the current flowing normally across the surface, as the surface becomes infinitesimally small. The current density represents the current flow per unit area at the point. In mediums that obey Ohm's law, the current density is proportional to the electric field intensity.

If s is the area surrounding the specified point at which the current flows, I is the magnitude of the current flowing, θ is the angle between the direction of current flow and the normal to the surface through which it flows, then the normal component of current across the surface is $I_n = I \cos \theta$, and the current density is defined to be

$$j = \lim_{\Delta s \rightarrow 0} \frac{\Delta I_n}{\Delta s}$$

In the mks system of units, current is measured in amperes, area is measured in square meters, and current density is, therefore, measured in amperes per square meter.

In terms of the current density, the total current flowing across a surface, s , is

$$I = \int_s j \cos \theta \, ds = \int_s j_n \, ds = \int_s j \cdot ds$$

If s_c is the conduction (and convection) component of the current density, and $s_d = K'(d\mathcal{D}/dt)$ is the displacement component, the total current density is then $s_T = s_c + s_d = s_c + K'(d\mathcal{D}/dt)$, and the normal component, flowing perpendicularly through a surface, is given by

$$s_n = \left(s_c + K' \frac{d\mathcal{D}}{dt} \right)_n$$

where, in the mks system, s_n is measured in amperes per square meter.

59. Continuity of Current. From the fact that elementary charges are indestructible, we may conclude then an electric current is continuous and that electricity is conserved. Such is indeed the case. The proof of this statement is readily available.¹

60. Magnetic Field Produced by Electric Currents. Ampère's experiments on the magnetic intensity produced by an electric current showed that the magnetic field intensity produced at a point is proportional to the number of conductors, N , the current, I , in each conductor, the length, l , of the current-carrying elements, and the sine of the angle, θ , between the direction of the current element and the radius vector, $r = OP$, from the current carrying element to the point P ; it is also inversely proportional to the square of the distance, r , between the current-carrying element and the point of observation. Mathematically,

$$\mathcal{H} = \frac{NI l \sin \theta}{r^2}$$

where, in mks units, the magnetic field intensity, \mathcal{H} , is measured in ampere-turns per meter. From the relation, $\mathcal{B} = \mu\mathcal{H}$, it follows that the magnetic flux density due to the flow of current through current-carrying elements is given by

$$\mathcal{B} = \frac{\mu(NI l \sin \theta)}{r^2}$$

where \mathcal{B} is measured in webers per square meter in mks units.

Lines of magnetic intensity surrounding a current-carrying conductor are represented in the diagram of Fig. 23.

61. Divergence of Magnetic Flux. For a magnetic field produced by the flow of an electric current, as well as in the magnetostatic case, the magnetic flux is given by

$$\phi = \int_s \mathcal{B} \cos \theta \, ds$$

¹ See, for example, such references as the following: R. W. P. King, "Electromagnetic Engineering," Vol. 1, McGraw-Hill, 1945; J. A. Stratton, "Electromagnetic Theory," McGraw-Hill, 1941; W. R. Smythe, "Static and Dynamic Electricity," McGraw-Hill, 1939.

where θ is the angle between the direction of \mathcal{B} and the normal to the surface element, ds . Over a closed surface,

$$\phi = \oint \mathcal{B} \cos \theta \, ds$$

i.e., the divergence of the magnetic flux, over a closed surface, is zero. This is represented diagrammatically in Fig. 23 by the illustration showing as many lines of magnetic intensity entering as leaving a small volume containing the current-carrying element.

62. Magnetic Force Produced by Current Elements. If a conductor of N turns, through which current I flows, lies in a magnetic field whose flux density is \mathcal{B} webers per square meter, the force exerted on the conductor is given by the relation

$$f = \mathcal{B}NI \sin \theta$$

where θ is the angle between the direction of \mathcal{B} and that of l .

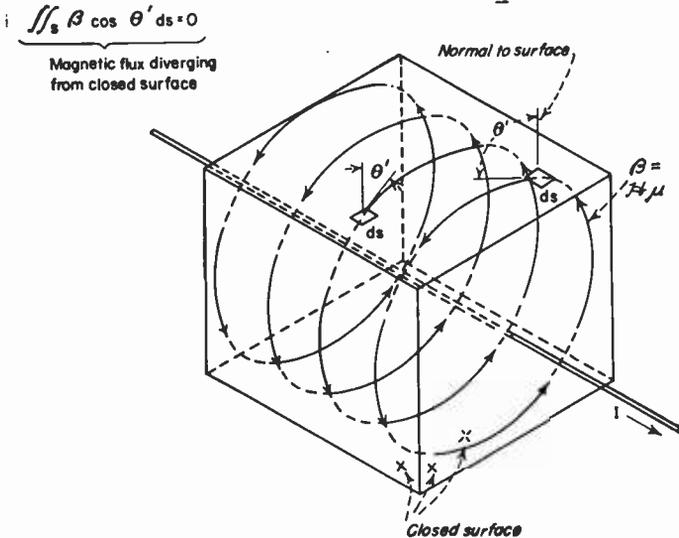


FIG. 23. Lines of magnetic intensity around a current-carrying conductor are continuous, and the total magnetic flux crossing a closed surface is zero.

Since a magnetic field can be produced by the flow of current through conductors, and further since a current-carrying element in a magnetic field has a force exerted upon it, it follows that a mechanical force must exist between two current-carrying conductors.

63. Electrical Conduction. The electric field intensity, \mathcal{E} , is the force that tends to move electric charges. Under the influence of such a force, a current density, \mathcal{J} , tends to flow in a conductor, the current density being represented by the time rate of flow of charges across an area perpendicular to the electric field intensity. The electric field intensity may then be viewed as the cause that produces or tends to produce a flow of conduction current density. There must, therefore, be some relation between \mathcal{E} and \mathcal{J} to express this cause-effect relation and, from what has been said about dielectrics and conductors, we might expect the results for these two cases to differ.

The relation between the electric field intensity and the current density cannot be obtained by simple reasoning alone; recourse must be made to experiment as well as

to theory to obtain this relationship. Experiments show, however, that for homogeneous solids, this relationship is given by

$$j = \sigma \mathcal{E}$$

where σ is a constant, characteristic of the material through which the electric field is established and through which the current flows, called the "conductivity" of the material. We may also obtain the correlation by means of the expression,

$$\mathcal{E} = \rho j$$

where $\rho = 1/\sigma$ is a characteristic of the material, known as its "resistivity." It should be observed that although the Greek letter sigma (σ) has been used for surface charge density as well as for conductivity, there is no connection between the two concepts expressed by the same symbol. Likewise, there is no connection between volume charge density and resistivity although both are expressed by the Greek rho (ρ). Little confusion will result because of the use of common symbols for different ideas, since the context will indicate which meaning is intended.

64. Conduction in Metals. Being an inherent property of a material to which an electric field is applied, the conductivity may be expected to be a function of the physi-

Table 7. Resistivity and Temperature Coefficient of Resistivity of Common Electrical Conductors at $t_0 = 20^\circ\text{C}$

Material	Resistivity, ohm-m	Temperature coefficient, A/ $^\circ\text{C}$
Aluminum.....	2.824×10^{-8}	0.0038
Antimony.....	41.7×10^{-8}	0.0036
Bismuth.....	120×10^{-8}	0.004
Brass.....	7×10^{-8}	0.001-0.002
Carbon.....	2.900×10^{-8}	-0.0005
Copper, annealed.....	1.724×10^{-8}	0.00393
Copper, hard-drawn.....	1.771×10^{-8}	0.00382
Gold.....	2.44×10^{-8}	0.0034
Iron.....	10.0×10^{-8}	0.0050
Mercury.....	95.783×10^{-8}	0.00089
Molybdenum.....	5.7×10^{-8}	0.0033
Nickel.....	7.8×10^{-8}	0.006
Platinum.....	10.0×10^{-8}	0.003
Silver.....	1.59×10^{-8}	0.0038
Tantalum.....	15.5×10^{-8}	0.00347
Thorium.....	40.1×10^{-8}	0.0021
Tungsten.....	5.6×10^{-8}	0.0045
Zinc.....	5.8×10^{-8}	0.00415

cal or chemical composition of the material, particularly with regard to those factors which determine or influence the relative freedom of motion of electrons within the material. This is, indeed, found to be true, so that the electrical conductivity or resistivity of a medium can be related to its characteristic atomic structure.

An elementary theory of metallic conduction¹ shows that the conductivity of a substance is given by the relation

$$\sigma = \frac{1}{2} \frac{l_m}{v_m} N e^2$$

where l_m is the mean free path of the electrons in the conductor, v_m is the average random velocity with which the electrons travel, N is the number of electrons in a length, l , of conductor of uniform cross section, and e is the charge of the electron. In mks units the conductivity is measured in mhos per meter. The resistivity, $\rho = 1/\sigma$, is measured in ohm-meters in the mks system of units.

¹ Page, L., and N. I. Adams, "Principles of Electricity," p. 162, Van Nostrand, 1931.

The number of free electrons, N , in any given length of uniform conductor may be expected to be a function of temperature, so that the electrical properties of conductors are likewise a function of temperature. For metallic conductors, the resistivity increases with the temperature, but there are some materials, of which carbon is a notable example, in which the resistivity decreases with an increase of temperature. The resistivity of a homogeneous material may be expressed, as a function of temperature, by the expression,

$$\rho = \rho_0[1 + A(t - t_0)]$$

where ρ_0 is the resistivity of the material at the specified or reference temperature, t_0 , t is a temperature at which the resistivity is desired, and A is a constant, called the "temperature coefficient" of resistivity, typical values for which are given in Table 7.

65. Force between Current-carrying Elements. It has been shown (Sec. 60) that a current-carrying circuit element establishes a magnetic field and (Sec. 62) that a force is exerted on a conductor carrying a current in a magnetic field. Therefore, if two conductor elements each carry an electric current, each will be in the magnetic field produced by the other; hence, a force will be exerted between the two. The force between these two elements is given by

$$f = K_{em} K' \left(\frac{\mu}{r^2} \right) (N_1 I_1 l_1 \sin \theta_1) (N_2 I_2 l_2 \sin \theta_2)$$

where N_1 and N_2 are the number of turns of the first and second circuit elements, I_1 and I_2 are the currents in the first and second elements, respectively, l_1 and l_2 are the lengths of the first and second conducting elements, and θ_1 and θ_2 are the angles between straight lines connecting the two current-carrying elements and the direction of the first and second elements, respectively. In mks units, f is measured in newtons, I in amperes, l in meters, and μ in henrys per meter. The factor K_{em} relates units of mechanical force with units of current, and K' is a factor, as given in Table 6, whose value depends on whether rationalized or unrationalized system of units are employed.

66. Relation between Static and Dynamic Force Fields. Coulomb's law for electric charges and magnetic poles provides a means of relating mechanical forces with the magnitudes of electric charges and magnetic poles. On the basis of Coulomb's law for electric charges, the electrostatic system of units has been developed. Quite independently, the magnetic system of units has been derived on the basis of Coulomb's law for magnetic poles. The force between two current-carrying elements provides a means for relating the static and dynamic force fields, since currents produce magnetic effects and are but the flow of charges, it also provides a means for checking and correlating the system of units established for the electrostatic and magnetostatic cases.

Equating the forces obtained from charges, single-turn current-carrying circuit elements, and magnetic poles, the three force expressions are,

$$f = K_e K' \frac{Q_1 Q_2}{r^2} = K_{em} K' \frac{1}{\mu r^2} (\mu I_1 l_1 \sin \theta_1) (\mu I_2 l_2 \sin \theta_2) = K_m K' \frac{M_1 M_2}{\mu r^2}$$

By the definition of unit current and magnetic shell, the last two expressions must be identical so that $K_{em} = K_m$. To simplify calculations, let $Q_1 = Q_2 = Q$, $M_1 = M_2 = M$, $I_1 l_1 = I_2 l_2 = Il$, and let $\theta_1 = \theta_2 = \theta = 90^\circ$ so that $\sin \theta = 1$. Then, for r the same in all cases

$$K_e \left(\frac{Q^2}{\epsilon} \right) = K_m \left(\frac{M^2}{\mu} \right) = K_m \mu (Il)^2$$

For a steady current, $I = Q/t$; hence,

$$K_e \left(\frac{Q^2}{\epsilon} \right) = K_m \mu \left(\frac{Ql}{t} \right)^2$$

from which

$$\frac{K_e}{K_m} \frac{1}{\mu\epsilon} = \left(\frac{l}{i}\right)^2 = v^2$$

where the velocity v is given by l/i .

If the permittivity and permeability of free space are both taken as unity, as they are in the cgs electrostatic and electromagnetic (magnetostatic) systems of units, the ratio of units in the two systems has the dimensions of velocity squared. Further examination of the units of measurement in these two systems shows that they are the same as the velocity of light in free space to well within the limits of experimental error. For mediums other than that of free space, the actual values of μ and ϵ must be used to determine the velocity of propagation of electrical effects.

The somewhat unexpected result that the ratio of dimensions in the cgs electrostatic and magnetostatic systems has the dimensions of velocity squared comes from the fact that values for permeability and permittivity were arbitrarily assigned at a time before there was any known relation between electricity and magnetism.

67. Magnetomotive Force. The magnetomotive force acting in any closed path in a magnetic field is the line integral of the magnetizing force around the path. In any closed path, the magnetomotive force resulting from the flow of current through a path is proportional to the current that links its path. In mks units, the magnetomotive force is expressed in ampere-turns, and around any closed path has the value

$$\oint \mathcal{H} \cos \theta \, dl = K'''NI$$

where θ is the angle between the direction of \mathcal{H} and that of l , N is the number of times the path links the magnetic field, I is the current in amperes, and K''' has the value given in Table 6.

68. Ampère's Circuital Law for Magnetic Intensity. In radio practice we are interested not only in conduction currents and circuit phenomena (in which current flows through filamentary circuit paths), but also in the general flow of current through space, more conveniently expressed in terms of current density. Furthermore, the current density may have displacement as well as conduction components. For the case in which current density flows through a surface, Ampère's circuital law of magnetic intensity may be stated as follows: The magnetizing force around a closed loop is equal to the total current flowing through the surface bounded by the loop. Mathematically,

$$\oint \mathcal{H} \cos \theta \, dl = \int \int_S K''' \left(s_c + K'' \frac{d\mathcal{D}}{dt} \right)_n \, ds = K''' \left(I_c + K'' \frac{d\psi}{dt} \right)$$

In this case s_c is the conduction component of the current density and $d\mathcal{D}/dt$ is the displacement component, while I_c is the conduction component of the total current and $d\psi/dt$ is the displacement component of current. The subscript, n , designates the normal component. In the mks system of units, current is measured in amperes, current density in amperes per square meter, and magnetic field intensity in ampere-turns per meter.

The integral form of Ampère's circuital law, above, is particularly suitable when the net over-all, or integrated effects of the resultant magnetic field and electric currents are of primary concern, as in dealing with electric circuits. When dealing with radiation phenomena, field concepts, or other matters in which the spatial distribution of electromagnetic effects is important, it is more convenient to deal with the differential form of Ampère's circuital equation in which the directional components of the current density and electric flux density are expressed in terms of the curl or rotational components of the magnetic field intensity.

In deriving the differential form of Ampère's circuital law, the resultant current flow is resolved into three orthogonal components; likewise the resulting magnetic field is resolved into three orthogonal components such that the circulation of the magnetic field is in a plane normal to the direction of flow of the electric current.

The resultant effect, at a point in space, is then the vector sum of the three orthogonal components.

In rectangular coordinates the differential expression for Ampère's circuital law takes the form

$$iK''' \left(g_x + K'' \frac{dD_x}{dt} \right) + jK''' \left(g_y + K'' \frac{dD_y}{dt} \right) + kK''' \left(g_z + K'' \frac{dD_z}{dt} \right) \\ = i \left(\frac{\partial C_x}{\partial y} - \frac{\partial C_y}{\partial x} \right) + j \left(\frac{\partial C_z}{\partial z} - \frac{\partial C_z}{\partial x} \right) + k \left(\frac{\partial C_y}{\partial x} - \frac{\partial C_x}{\partial y} \right)$$

The sum of the three directional components gives the relation, at a point, for Ampère's circuital law.

The diagram of Fig. 24 represents the relation existing between the direction of the flow of current, the direction of the magnetic field intensity, and the direction of a closed path, as called for by Ampère's circuital law.

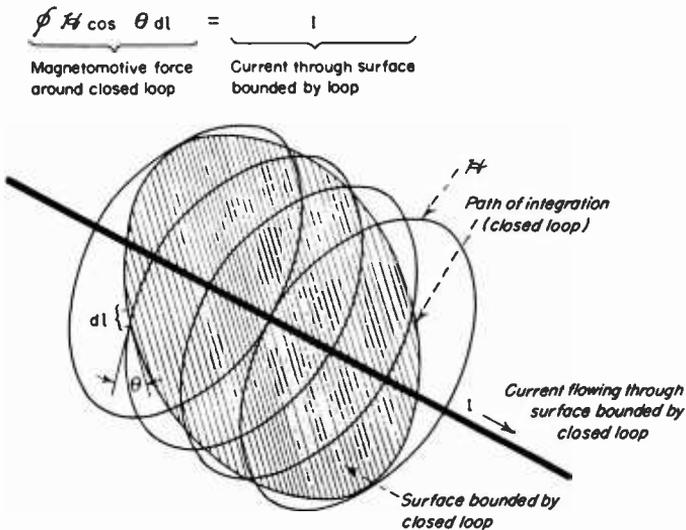


FIG. 24. Diagram representing the integral form of Ampère's circuital law.

69. Electromotive Force. Electromotive force (emf) is that property of a physical device which tends to produce an electric current in a circuit. It is also that property of a system which tends to alter the motion of electricity, or to maintain its motion against resistance. Like the difference of potential, emf is measured by the energy per unit charge imparted to the charge as it travels through a region in which the emf acts. It is evident, therefore, that the emf is not, as its name would imply, a force, but rather work per unit charge. In the mks system of units, the unit of emf is the joule per coulomb, to which the name "volt" has been given.

One form of emf has already been encountered, that of electrostatic potential difference, which exists by virtue of the position of electric charges. Sources of electric energy, such as batteries, thermocouples, and photoelectric devices, are capable of producing an emf by chemical action or by the action of radiant energy. Another type of emf, first discovered by Faraday, results from the motion of a conductor in a magnetic field; it is called the "induced" emf, or the "electromotive force of induction."

70. Faraday's Circuital Law for Induced Emf. As a result of extensive researches, Faraday, in 1833, was able to announce his discovery of the relation governing an

emf appearing across a section of a conductor having motion relative to a magnetic field in which it is situated. In the following year, Lenz clearly specified the relationships between the directions of the magnetic field, the motion of the conductor relative to the magnetic field, and the polarity or direction of the resulting emf. As amplified by Lenz, Faraday's law may be stated as follows: When the change in magnetic flux linkages is caused by the relative motion of a magnetic field and a conductor forming part of an electric circuit, an electromotive force is induced in the circuit, the magnitude of which is equal to the time rate of decrease of the flux linkages.

A flux linkage is defined as the interlinking of a line of magnetic flux with a single conductor, and is represented by the symbol Λ . Therefore, $\Lambda = N\phi$ flux linkages signifies the product of ϕ lines of magnetic flux linking N turns of a conductor or, conversely, N conductors linking the magnetic flux, ϕ . With this notation, the magnitude and direction of the emf induced in a conductor is given by

$$E = - \frac{d\Lambda}{dt} = - \frac{d(N\phi)}{dt}$$

where, in mks units, E is the induced emf in volts, N is the number of conductors (or number of turns of a single conductor closely bunched together), and ϕ is the magnetic flux in webers.

The relation given above is quite suitable for electric circuits but, when dealing with field phenomena, the equations are more convenient when expressed in terms of the electric field intensity and the magnetic flux density together with elements of length and surface. Since the drop in emf around a closed loop is given by the relation $\oint \epsilon \cos \theta \, dl$ and the magnetic flux traversing the surface enclosed by the loop is given by $\iint_s \mathfrak{B} \cos \theta' \, ds$, Faraday's law for electromotive induction may be written as

$$\oint \epsilon \cos \theta \, dl = - \frac{\partial}{\partial t} \iint_s \mathfrak{B} \cos \theta' \, ds$$

where (in mks units) ϵ is the electric field intensity given in volts per meter, \mathfrak{B} is the magnetic flux density in webers per square meter, dl is an element of length in meters, ds is an element of area in square meters, and θ is the angle between the direction of the electric field intensity and the line element, whereas θ' is the angle between the direction of the magnetic flux density and the normal to the surface through which the flux passes.

The magnetic flux density is a continuous function of time, t , and surface, s . For such conditions it is permissible to perform the differentiation under the integral sign¹ so that Faraday's circuital law may also be written in the form

$$\oint \epsilon \cos \theta \, dl = - \iint_s \frac{\partial(\mathfrak{B} \cos \theta')}{\partial t} \, ds$$

where the symbolic notation is the same as before.

At a point, this field relation may be expressed in the differential form

$$i \left(\frac{\partial \epsilon_x}{\partial y} - \frac{\partial \epsilon_y}{\partial z} \right) + j \left(\frac{\partial \epsilon_x}{\partial z} - \frac{\partial \epsilon_z}{\partial x} \right) + k \left(\frac{\partial \epsilon_y}{\partial x} - \frac{\partial \epsilon_x}{\partial y} \right) = - \left(i \frac{\partial \mathfrak{B}_z}{\partial t} + j \frac{\partial \mathfrak{B}_y}{\partial t} + k \frac{\partial \mathfrak{B}_x}{\partial t} \right)$$

where, in mks units, ϵ is the electric field intensity in volts per meter, \mathfrak{B} is the magnetic flux density in webers per square meter, and the directional components are indicated in rectangular coordinates, by the subscripts.

In the diagram of Fig. 25 is represented flux density, \mathfrak{B} , flowing through the surface, s , enclosed by a looped conductor. The direction between the plane of the closed conductor and the plane normal to that of the magnetic flux is θ , and the direction or polarity of the emf induced in the loop is also indicated.

¹ Sokolnikoff, I. S., and E. S. Sokolnikoff, "Higher Mathematics for Engineers and Physicists," McGraw-Hill, 1941.

71. Potential Difference, Electromotive Force, and Voltage. The terms "potential difference," "electromotive force," and "voltage" are often used loosely and indiscriminately, although there is a clear distinction between the first two of these at least.

Strictly speaking, the potential difference is the difference in electrostatic potential due to the distribution of electric charges, and the magnitude of the potential difference depends upon the magnitude and positions of the charges. The concept is also extended, however, to the case of moving charges, *i.e.*, the case in which current flows, in which case the potential difference is interpreted, at any instant of time, to be the same as the electrostatic potential difference for the charge configuration at the instant in question. This extension of the concept of electrostatic potential difference is valid for time variations of the charges, or frequencies of current, so long as the electromagnetic field is essentially of equal magnitude, at any instant of time, for

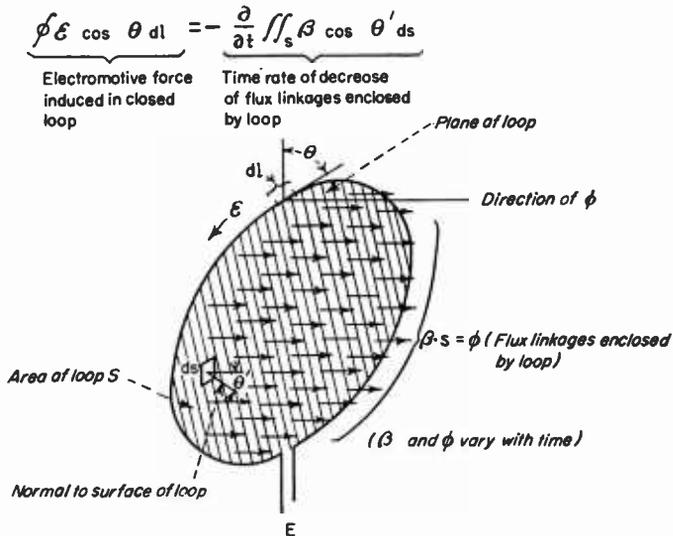


FIG. 25. Representation of integral form of Faraday's circuital law for induced emf.

the entire region occupied by the electrical system with which it is associated. At sufficiently high frequencies, the concept is no longer applicable.

An electric field may be produced by the distribution of electric charges or, in accordance with Faraday's circuital law, it may be produced by the relative motion of a conductor and a magnetic field. The latter case, that of the electrodynamic field, is distinguished from the electrostatic field in not producing a potential difference in the absence of material substance, but a conductor in a field of induction has a potential difference across it.

The term "voltage" is used to designate potential difference whether due to the simple or extended concepts of electrostatic potential or due to induced emf.

It is worth noting that currents are not necessarily induced. If emf is induced in a closed circuit, a conduction current will flow, but the emf will be induced whether the circuit is open or not. Conduction current cannot flow in an open circuit, however.

As mentioned in Sec. 37, if A and B are the two points between which a positive charge is transferred, then B is at a higher potential than A if external work is done in transferring the charge from A to B , *i.e.*, if energy is expended to effect this transfer. Point B is then said to be at a higher potential than point A ; alternatively, point B is said to be positive with respect to point A . Energy is given up when the positive charge is transferred from B to A .

References

- Abraham, M., and R. Becker: "Classical Theory of Electricity and Magnetism," Blackie, 1932.
- Bennett, E., and H. M. Crothers: "Introductory Electrodynamics for Engineers," McGraw-Hill, 1926.
- Culliwiek, E. G.: "Fundamentals of Electromagnetism," Cambridge, 1939.
- Dampier-Whetham, W. C.: "Theory of Experimental Electricity," Cambridge, 1923.
- Faraday, M.: "Experimental Researches in Electricity," Dutton, 1931.
- Gray, A.: "Treatise on Magnetism and Electricity," Methuen, 1901.
- Grimsehl, E., and R. Tomaschek: "Electricity and Magnetism," London-Glasgow, 1933.
- Hirst, A. W.: "Electricity and Magnetism for Engineering Students," Prentice-Hall, 1937.
- Jeans, J. H.: "Mathematical Theory of Electricity and Magnetism," Cambridge, 1925.
- Kelly, H. C.: "A Textbook of Electricity and Magnetism," Wiley, 1941.
- Lawrence, R. R.: "Principles of Alternating Currents," McGraw-Hill, 1935.
- Loeb, L. B.: "Fundamentals of Electricity and Magnetism," Wiley, 1931.
- Maxwell, J. C.: "Electricity and Magnetism," Oxford, 1881.
- Moullin, E. B.: "Principles of Electromagnetism," Oxford, 1932.
- Page, L.: "Introduction to Theoretical Physics," Van Nostrand, 1928.
- , and N. I. Adams: "Electrodynamics," Van Nostrand, 1940.
- , and ———: "Principles of Electricity," Van Nostrand, 1931.
- Pender, H.: "Electricity and Magnetism for Engineers," McGraw-Hill, 1918.
- , and S. R. Warren: "Electric Circuits and Fields," McGraw-Hill, 1943.
- Pidduck, F. B.: "Treatise on Electricity," Cambridge, 1925.
- Pilley, J.: "Electricity," Oxford, 1933.
- Planck, M.: "Theory of Electricity and Magnetism," Macmillan, 1932.
- Pohl, R. W.: "Physical Principles of Electricity and Magnetism," Blackie, 1930.
- Poor, V. C.: "Electricity and Magnetism," Wiley, 1931.
- Poynting, J. H., and J. J. Thomson: "Electricity and Magnetism," London, 1914.
- Ramsey, A. S.: "Electricity and Magnetism," Cambridge, 1937.
- Starling, S. G.: "Electricity and Magnetism for Advanced Students," Longmans, 1912.
- Suydam, V. A.: "Fundamentals of Electricity and Electromagnetism," Van Nostrand, 1940.
- Thomson, J. J.: "Elements of the Mathematical Theory of Electricity and Magnetism," Van Nostrand, 1921.
- Thomson, W.: "Papers on Electrostatics and Magnetism," Macmillan, 1884.
- Webster, A. G.: "Electricity and Magnetism," Macmillan, 1897.
- Wilson, W.: "Electromagnetism and Optics," London, 1933.

THE ELECTROMAGNETIC FIELD

72. Electromagnetic Field Relations. As concerns its electrical aspects, the basis of radio communication resides in the four electromagnetic field relations known as Maxwell's equations, which have already been developed and are summarized in Table 8. They are fundamental in the sense that from them can be derived relations of circuit behavior, the wave equation, the propagation characteristics of electromagnetic waves (if enough is known of the transmission medium), and the like.

Maxwell's electromagnetic field equations should be regarded as a compact and generalized means of expressing known electrical phenomena. This set of relationships differs from most equations used by engineers in that they do not provide information in a form that is directly applicable for design purposes. In fact, the use of these equations for most problems of engineering application usually involves such difficulties of specifying boundary conditions and mathematical manipulation that the Maxwell equations have little direct usefulness to engineers. But by no means do these limitations invalidate their utility, for this set of relations can be made to yield other relationships (*e.g.*, the laws for electric circuits) applicable in special cases and sufficiently easy to manipulate as to be of great practical utility.

In Table 8 each of the four Maxwell equations specifying a fundamental electromagnetic relation is interpreted in six different ways in addition to the pictorial representation referred to by the figure numbers. The first column gives the usual name of the relationship. The second column provides a word statement of each relation, which aims to express the fundamental physical behavior without recourse to mathematical notation. The next column gives the figure number in which the approximate physical relations are pictorialized. The next columns express the result of physical observations in the classical form of integral equations which take account of the total, net, or resultant of all directional components of the field. The last columns are the equivalent integral expression in the notation of vector analysis.

Table 8. Alternative Statements of Maxwell's Equations

Name	Statement	Fig.	Classical form		Vector form	
			Integral	Differential (point rel.)	Integral	Differential (point rel.)
1. Gauss's theorem for electric flux density (Coulomb's law for single di-electric)	The electric flux diverging from a closed surface, ψ , is equal to K times the total density enclosed by the surface or to K times the charges enclosed by the surface enclosing the volume	22	$\psi = \oiint_s \mathcal{D} \cos \theta' ds$ $= \iiint_v \rho d\tau$ $= K'''Q$	$\frac{\partial \mathcal{D}_x}{\partial x} + \frac{\partial \mathcal{D}_y}{\partial y} + \frac{\partial \mathcal{D}_z}{\partial z} = \begin{cases} K''' \rho \\ \text{or} \\ 0 \end{cases}$	$\psi = \oiint_s \mathcal{D} \cdot n ds$	$\text{div } \mathcal{D} = \nabla \cdot \mathcal{D} = \begin{cases} K''' \rho \\ \text{or} \\ 0 \end{cases}$
2. Gauss's theorem for magnetic flux density (Coulomb's law for a single magnetic medium, or law of magnetic flux density)	The magnetic flux diverging from a closed surface, ϕ , is zero. It follows that magnetic lines of force are continuous, or form closed loops	23	$\phi = \oiint_s \mathcal{B} \cos \theta' ds$ $= 0$	$\frac{\partial \mathcal{B}_x}{\partial x} + \frac{\partial \mathcal{B}_y}{\partial y} + \frac{\partial \mathcal{B}_z}{\partial z} = 0$	$\phi = \oiint_s \mathcal{B} \cdot ds$	$\text{div } \mathcal{B} = \nabla \cdot \mathcal{B} = 0$
3. Circuital law for magnetic intensity (from Ampere and Maxwell)	The magnetomotive force around a closed loop, \mathcal{F} , is proportional to the total current flowing through the surface bounded by the loop	24	$\mathcal{F} = \oint \mathcal{H} \cos \theta dl$ $= \iint_s K''' (g_x + K'' \frac{\partial \mathcal{D}_x}{\partial t}) + K' \frac{\partial \mathcal{D}}{\partial t} \cdot n ds$	$\frac{\partial \mathcal{H}_x}{\partial y} - \frac{\partial \mathcal{H}_y}{\partial x} = K''' (g_z + K'' \frac{\partial \mathcal{D}_z}{\partial t})$ $\frac{\partial \mathcal{H}_z}{\partial x} - \frac{\partial \mathcal{H}_x}{\partial z} = K''' (g_y + K'' \frac{\partial \mathcal{D}_y}{\partial t})$ $\frac{\partial \mathcal{H}_y}{\partial z} - \frac{\partial \mathcal{H}_z}{\partial y} = K''' (g_x + K'' \frac{\partial \mathcal{D}_x}{\partial t})$	$\mathcal{F} = \oint \mathcal{H} \cdot dl$ $= \iint_s K''' (g_x + K'' \frac{\partial \mathcal{D}_x}{\partial t}) + K'' \frac{\partial \mathcal{D}}{\partial t} \cdot ds$	$\text{curl } \mathcal{H} = \nabla \times \mathcal{H} = K''' (g_x + K'' \frac{\partial \mathcal{D}}{\partial t})$
4. Circuital law for induced emf (generalized law of Faraday)	In all cases in which the flux linkages change, the algebraic value of induced emf in a closed loop, E , is equal to the time rate of decrease of flux linkages enclosed by the loop	25	$E = \oint \mathcal{E} \cos \theta dl$ $= - \frac{\partial}{\partial t} \iint_s \mathcal{B} \cos \theta' ds$ $= - \iint_s \frac{\partial (\mathcal{B} \cos \theta')}{\partial t} ds$	$\frac{\partial \mathcal{E}_x}{\partial y} - \frac{\partial \mathcal{E}_y}{\partial x} = - \frac{\partial \mathcal{B}_z}{\partial t}$ $\frac{\partial \mathcal{E}_z}{\partial x} - \frac{\partial \mathcal{E}_x}{\partial z} = - \frac{\partial \mathcal{B}_y}{\partial t}$ $\frac{\partial \mathcal{E}_y}{\partial z} - \frac{\partial \mathcal{E}_z}{\partial y} = - \frac{\partial \mathcal{B}_x}{\partial t}$	$E = \oint \mathcal{E} \cdot dl$ $= - \frac{\partial}{\partial t} \iint_s \mathcal{B} \cdot ds$ $= - \iint_s \frac{\partial \mathcal{B}}{\partial t} \cdot ds$	$\text{curl } \mathcal{E} = \nabla \times \mathcal{E} = - \frac{\partial \mathcal{B}}{\partial t}$

The integral form is the basis of circuit equations and is most suitable where the integrated electric and magnetic fields are most conveniently expressed in terms of voltage and current. Each of the field components is also expressed in the form of a differential equation, one column giving the classical form (in rectangular coordinates only) and the other giving the equivalent in vector notation. As given in Table 8, the equations are suitable for either the rationalized or unrationalized system. The differential form of the field equations is most suitable for the determination of relations at a point, or where it is necessary to take into account the spatial and directional distributions of the electric and magnetic effects of electromagnetic phenomena. The classical method of writing these equations is somewhat long but, for purposes of physical interpretation, has the advantage of illustrating the manner in which each directional component of the electromagnetic field is separately involved. The vector form is much more compact and has the advantage that the same compact notation is used for any orthogonal system of coordinates. A working knowledge of vector analysis is required before the physical interpretation of the equations is evident, however.

73. Maxwell's Equations for Special Cases. Maxwell's equations may be grouped into five distinct divisions, depending upon whether or not certain effects are functions of time and whether the magnitude of certain physical effects is sufficiently small to be negligible. From an engineering point of view, the separate cases are (1) the static case, (2) the steady case for stationary electric and magnetic fields, (3) the quasi-steady case for closed circuits in which, except for capacitors, the displacement component of the current is sufficiently small to be negligible, (4) the quasi-steady case for open circuits in which displacement current is no longer negligible, and (5) the general case of dynamic phenomena in which the magnitude of the displacement current is appreciable and may exceed that of conduction current. These cases are discussed below.

1. *The Static Case.* The static case is that, encountered in electrostatics and magnetostatics, for which no time variations of electric charges or magnetic poles occur. There is no flow of charge and, therefore, there can be neither conduction nor displacement currents. Except for Gauss's theorem for electric charges, the right-hand side of all the Maxwell equations reduces to zero; it may even reduce to zero for Gauss's theorem for electric charge if the medium is charge-free. The electrostatic and magnetostatic fields have independent laws (the divergence equations deduced from Coulomb's law) even though the mathematical forms of these laws are similar. Energy is associated with the electrostatic field and also, separately, with the magnetostatic field. The electrostatic and magnetostatic fields become superimposed upon one another without influencing each other in the slightest degree; hence, the complete segregation of the two kinds of fields and the impossibility of determining the connection between electricity and magnetism for the static case.

Although the static case has been important in the development of the classical theory of electricity, it can obviously have no direct application in conveying electrical communication, for the essence of electrical communication is the time variation of some electrical quantity in accordance with the information to be transmitted.

2. *Steady Case for Stationary Electromagnetic Field.* The steady case is that, encountered in d-c circuits and systems, for which there is a constant and steady flow of electric charges to produce a conduction (or convection) current, but no displacement current is possible. All of Maxwell's equations have interpretation except that expressing Faraday's circuital law of induced emf which, of course, depends upon a time-varying magnetic field. Electric and magnetic effects are now no longer independent of one another. The electric field in a conductor establishes a streamline of flow of electric charges which, in turn, gives rise to a magnetic field that is constant with respect to time and does not react to modify the electric field. A continual transformation of electric energy into heat takes place. Lumped constant circuit elements can be rigorously defined for this case; in fact, it is the only case for which the circuit parameters are rigorously defined. Yet, in the true steady case, *i. e.*, free from tran-

sients, only the resistance parameter is encountered in circuit behavior; inductance and capacitance determine circuit behavior only for transient conditions.

The steady case—the d-c or stationary electromagnetic field case—has no direct active role in electrical communication that requires time variations of some electrical quantity. Nevertheless, the steady case is of considerable indirect interest to engineers because certain important concepts (*e.g.*, lumped circuit parameters) are rigorously defined only for this case. Moreover, steady voltages and currents are often required in order that certain pieces of radio equipment (*e.g.*, electron tubes) may operate properly in their communication application.

3. *Quasi-steady Case for Closed Circuits.* This is probably the best known of the five cases since it is the most frequently encountered in electrical engineering. For this case the conduction current density, the magnetic flux linkages, and the induced emf are all single-valued functions of time and (except in capacitors) the displacement current, although not zero, is negligible compared to the conduction (or convection) component and hence may be ignored or neglected. For this case consideration is restricted to electrical processes occurring so slowly that, at any instant of time, the entire electromagnetic field surrounding the electrical system may be regarded as uniform. The electric and magnetic effects for this case are mutually related to the extent that a varying current produces a varying magnetic field which is capable of inducing a varying emf in a neighboring conductor which establishes a varying current if the conductor is closed. The fact that the displacement current is negligible precludes the possibility of radiation, or the mutual self-sustenance of electric and magnetic field intensities in dielectrics. All Maxwell equations have interpretation, but (except in capacitors) displacement currents are negligibly small and, therefore, are omitted from consideration. At any instant of time the current at all points in a (series) circuit is the same. The magnitude of the displacement current in capacitors is equal to the magnitude of the conduction current in a series circuit. A very convenient and highly useful simplification of Maxwell's equation results from neglecting the displacement current.

The very great advantage obtained for this case is that the number of variables determining the state of the electromagnetic field is greatly reduced. In fact, for most problems in this case, it is possible to omit consideration of space variables and deal only with time variations of the voltage and current—the integrated effects of the electric and magnetic field components. In the case of linear conductors the analysis can be still further simplified, for the current is determined by the voltage; hence, the entire phenomena may be expressed in terms of a single varying time function—either the current or the voltage. Usually the applied voltage is known, and if the current is then determined as a function of time, the problem of electrical behavior for this case is considered to be solved.

As in the steady case, the electric field intensity establishes a streamline of flow (tangential along the conductors) of electric charges of varying magnitude producing a varying magnetic field about the conductor.

For this case, lumped circuit parameters are no longer rigorously defined because displacement current is not taken into account. Therefore, the use of lumped circuit parameters leads to approximate rather than exact solutions of the electric circuit behavior. The errors involved in such approximations are usually sufficiently small as to be completely negligible for engineering (and for a good deal of scientific) work. The magnitude of the displacement current relative to that of the conduction current—and, hence, the error due to circuit analysis based on steady-state derivations—increases with frequency. Such error becomes appreciable (usually for frequencies above 50 to 100 Mc) for those cases in which the dimensions of the electrical system are comparable to a wavelength of oscillation or (what is the same thing) for those cases for which the instantaneous electromagnetic field can no longer be regarded as uniform throughout the space occupied by the electrical system. Distributed circuit parameters, *i.e.*, those circuit parameters whose properties are spatially distributed throughout the region of the electrical system instead of being considered as lumped

at a point in space, may often be used in place of lumped circuit parameters to obtain solutions of electric circuits (such as transmission lines) in which at least one dimension is an appreciable portion of a wavelength or more.

4. *Quasi-steady Case for Open Circuits.* This case is identical to that above with the following exceptions: Because displacement currents are no longer of negligible magnitude, the current at all points of an open circuit is no longer the same as in the case for closed circuits. The flow of energy is no longer tangential along conductors but has a perpendicular component as well. The electric energy of the field may no longer be neglected in comparison with the magnetic energy, and the magnetic energy can no longer be expressed accurately by conditions that apply for stationary currents. Radiation phenomena begin to become important, for energy may leave the open circuit and be propagated into space. These conditions are a result of the fact that the displacement current can no longer be regarded as negligible, although for this case it is not yet regarded as exceeding the magnitude of the conduction current.

5. *General Case for Dynamic Phenomena.* For this case, no restrictions are placed on Maxwell's equations. It is usually exceedingly difficult to apply the general equations to engineering problems, and this practical difficulty impedes the application of the general equations. For this case, the displacement current has appreciable magnitude. All four Maxwell equations (and all terms in each equation) are significant and must be used in their complete form. Circuit concepts are, generally, no longer valid. For this case, the electromagnetic field can no longer be regarded as uniform throughout the spatial distribution of the electrical system at any instant of time. Consequently the general case must be employed when very high frequencies are encountered or when the dimensions of the electrical system are comparable to a wavelength of oscillation.

Because displacement currents are appreciable, there is a mutual and self-sustained reaction between the electric and magnetic components of the electromagnetic field that leads to the phenomena of radiation.

74. *Circuit and Field Phenomena.* For convenience in radio engineering practice, Maxwell's equations may be divided into two broad categories in which the quasi-steady case for closed circuits and the general case are the most frequently employed.

Table 9. Comparison of Electromagnetic Field and Electric Circuit Concepts¹

Electromagnetic Field Concept	Electric Circuit Concept
Electric field intensity at a point in the medium	Electric voltage in branches of an electric circuit
Magnetic field intensity at a point in the medium	Electric current flowing through an electric circuit branch
Resistivity, or resistance per unit cube, of the medium	Resistance in a branch of the circuit
Permeability of the medium	Inductance of a branch of the circuit
Permittivity of the medium	Capacitance of a branch of the circuit
Energy stored in the electric field of unit volume of the medium	Energy stored in the dielectric of a capacitor
Energy stored in the magnetic field for unit volume of the medium	Energy stored in the magnetic field about an inductor
Energy dissipated in unit volume of the medium	Energy dissipated as heat in a conductor or resistor

¹ This table is a slight modification of that appearing in J. G. Brainerd, G. Koehler, H. J. Reich, and L. F. Woodruff, "Ultra-high-frequency Techniques," p. 371, Van Nostrand, 1942.

Upon the first of these cases is built the highly convenient theory using lumped circuit parameters, which are usually but not necessarily considered to be linear. The electromagnetic wave theory (including radiation phenomena and the behavior of electrical systems at very high frequencies) must necessarily be built upon the more general case of Maxwell's equations. The points of view in these two cases differ appreciably, as does also the emphasis with respect to spatial considerations. The approximately analogous concepts for these two cases are given in Table 9, which is intended to draw attention to the change in emphasis required for the two cases rather than to establish rigorous parallelisms.

The relations between circuit and electromagnetic field theory have been ably stated by Carson¹ in the following words:

The equations of electric circuit theory in their present form are essentially a generalization of the observations of Ohm, Faraday, Henry, Kirchhoff, and others, and their development preceded the electromagnetic theory of Maxwell and Lorentz. Naturally, in view of its early development, circuit theory embodies approximations, the precision of which cannot be determined from the observations on which it is based. For example, circuit theory explicitly ignores the finite velocity of propagation of electromagnetic disturbances, and hence the phenomena of radiation. Again it involves the assumption that the network can be represented by a finite number of coordinates and, thus, that it constitutes a rigid dynamic system. The rigorous equations of electromagnetic theory formulate the relations between current and charge densities and the accompanying fields. Circuit theory, on the other hand, expresses approximate relations between total currents and impressed electromotive forces.

References

- Abraham, M., and R. Becker: "The Classical Theory of Electricity and Magnetism," Blackie, 1932.
 Atwood, S. S.: "Electric and Magnetic Fields," Wiley, 1932.
 Barnett, S. J.: "Elements of Electromagnetic Theory," New York, 1903.
 Bateman, H.: "Mathematical Analysis of Electrical and Optical Wave Motion on the Basis of Maxwell's Equations," Cambridge, 1915.
 Brainerd, J. G., G. Koehler, H. J. Reich, and L. F. Woodruff: "Ultra-high-frequency Techniques," Van Nostrand, 1942.
 Briggs, H. F.: "The Electromagnetic Field," Clarendon Press, 1934.
 Carson, J. R.: "Electromagnetic Theory and the Foundations of Electric Circuit Theory," *Bell System Tech. J.*, 6 (1), 1, January, 1927.
 Duff, A. W., and S. J. Plimpton: "Elements of Electromagnetic Theory," Blakiston, 1940.
 Einstein, A.: On the Electrodynamics of Moving Bodies (from "Principles of Relativity") English trans., 1923.
 Frank, N. H.: "Introduction to Electricity and Optics," McGraw-Hill, 1940.
 Hague, B.: "Electromagnetic Problems in Electrical Engineering," Oxford, 1932.
 Harnwell, G. P.: "Principles of Electricity and Magnetism," McGraw-Hill, 1938.
 Heaviside, O.: "Electromagnetic Theory," 3 vols., London, 1893.
 Hertz, H.: "Electric Waves," London, 1893.
 Hund, A.: "Phenomena in High-frequency Systems," McGraw-Hill, 1936.
 Jeans, J. H.: "Mathematical Theory of Electricity and Magnetism," Cambridge, 1943.
 King, R.: "Electromagnetic Engineering," Vol. 1, McGraw-Hill, 1945.
 Larmor, J.: "Aether and Matter," Cambridge, 1900.
 Livens, G. H.: "Theory of Electricity," Cambridge, 1926.
 Mason, M., and W. Weaver: "The Electromagnetic Field," University of Chicago Press, 1929.
 Maxwell, J. C.: "Treatise on Electricity and Magnetism," Oxford, 1904.
 Moullin, E. B.: "Principles of Electromagnetism," Oxford, 1932.
 O'Rahilly, A.: "Electromagnetics—A Discussion of Fundamentals," Longmans, 1938. (Contains long list of references.)
 Page, L.: "Introduction to Electrodynamics," Boston, 1922.
 Pender, H., and K. McIlwain: "High Frequency Alternating Currents," Wiley, 1931.
 ———, and S. R. Warren: "Electric Circuits and Fields," McGraw-Hill, 1943.
 Pierce, G. W.: "Electric Oscillations and Electric Waves," McGraw-Hill, 1920.
 Planck, M.: "Theory of Electricity and Magnetism," Macmillan, 1932.
 Ramo, S., and J. R. Whinnery: "Fields and Waves in Modern Radio," Wiley, 1944.
 Rowland, H. J., On the Magnetic Effect of Electric Convection, *Am. J. Sci.*, 16, 30 (1878).
 Sarbacher, R. A., and W. Edson: "Hyper and Ultra High Frequency Engineering," Wiley, 1943.
 Schelkunoff, S. A.: "Electromagnetic Waves," Van Nostrand, 1943.
 Schott, G. A.: "Electromagnetic Radiation," Cambridge, 1912.
 Selgin, P. J.: "Electrical Transmission in Steady State," McGraw-Hill, 1946.
 Skilling, H. H.: "Fundamentals of Electric Waves," Wiley, 1942.
 Slater, J. C.: "Microwave Transmission," McGraw-Hill, 1942.
 Smythe, W. R.: "Static and Dynamic Electricity," McGraw-Hill, 1939.
 Stratton, J. A.: "Electromagnetic Theory," McGraw-Hill, 1941.
 Thomson, J. J.: "Elements of the Mathematical Theory of Electricity and Magnetism," Cambridge, 1904.
 Webster, A. G.: "Theory of Electricity and Magnetism," Macmillan, 1897.
 White, F.: "Electromagnetic Waves," London, 1934.

ELECTRIC CIRCUITS

75. Electric Circuit Concepts. An electric circuit is a path or a group of interconnected paths capable of carrying conduction electric currents. A closed electric

¹Carson, John R., "Electromagnetic Theory and the Foundations of Electric Circuit Theory," *Bell System Tech. J.*, 6 (1), 1, January, 1927.

circuit is a continuous path in the form of a loop or group of interconnected loops in which each loop is capable of carrying an electric current.

An assembly of lumped circuit elements is an electric circuit in which dissipation of energy and the storage of electric and magnetic energy may be considered as concentrated separately in each of three different circuit elements of physically small size. The physical size of lumped circuit elements is sufficiently small (theoretically infinitesimal) that, at any instant of time and over the region of space occupied by the circuit, the electric and magnetic fields are uniform. This condition is rigorously true for the steady (d-c) case and very nearly true if time variations of the electric and magnetic field are sufficiently slow. A circuit element is a physical device (usually with two terminals) exhibiting the electrical property of (1) energy dissipation or conversion of electrical energy into heat in a resistor, (2) storage of electric energy in a capacitor (or condenser), or (3) storage of magnetic energy in an inductor, or coil. These three circuit elements dissipate or store energy, and a drop of potential difference exists across them when an electric current flows through them. A fourth kind of circuit element, called a "source" of electric energy, transforms chemical, mechanical, or other energy into electric energy and has a rise of potential across its potentials.

A linear circuit is an electric circuit in which a linear differential equation precisely expresses the relations between the magnitudes of the instantaneous current and that of the rises and drops of potential difference across the circuit elements. A linear circuit is composed of circuit elements whose parameters (whose properties for storing or dissipating energy) are independent of the magnitude of the voltage across, or the current through, the elements. Those circuit elements whose properties are dependent upon the voltage or current are termed "nonlinear" circuit elements.

Lumped or concentrated circuit elements provide a convenient means for dealing with the integrated effects of electrical phenomena since (except in the case of capacitors in which electric energy is stored) only conduction current is important. The magnitude of the displacement current in capacitors is equal to the conduction current in the leads to the capacitor; hence, circuit phenomena are completely expressible in terms of the conduction current.

In developing the concepts of lumped circuit elements, all electrical effects, including the electric and magnetic fields, taking place are limited to the immediate vicinity of the electrical system or circuit. A source of electric energy (rise of potential difference) establishes electric field intensity along the physical configuration of the circuit elements, and all currents flow along the circuit conductors as a streamline, in the same direction as that of the electric field intensity, or the drop in potential difference.¹ The geometric configuration of the circuit, therefore, determines the space geometry of the system. Accordingly, it is possible to dispense with considerations of space coordinates in lumped circuit phenomena and consider only the time variations of the integrated electric field intensity and current density, *i.e.*, the time variations of the potential differences and the current. The concept of lumped circuit elements permits the solution—even if only approximate—of many problems that are beyond our present ability to solve by means of the more rigorous method employing electromagnetic field relations.

Strictly speaking, lumped circuit concepts apply only to the case in which the electric field intensity and current density are invariant with time, *i.e.*, to the steady electromagnetic field or the d-c case. As a practical matter, with suitable modifications or extensions, the steady-case concepts are applicable to those cases in which time variations of \mathcal{E} and \mathcal{J} (or E and I) are present, provided the dimensions of the circuit are sufficiently small so that, at any instant of time, the electric and magnetic fields are essentially uniform throughout the region occupied by the circuit and provided that (except in capacitors) displacement currents are negligible and, therefore, radiation is inappreciable.

¹ This statement applies to the conventional direction of the current; the electron flow is opposite to that stated here.

The idea of lumped circuit elements applies to conducting circuits composed of filamentary conductors through which the current flows and along which there are differences of potential or of electric field intensity. An exception to this statement must be made for capacitors, for the capacitance parameter necessarily involves a volume, across opposite faces of which exists a difference of potential.

Circuit phenomena deal with the integrated effects of \mathcal{E} and \mathcal{I} ; hence, these two quantities may be replaced by potential difference (E) and current (I), respectively. In developing the concept of lumped circuit constants, use is made of the relations

$$\begin{aligned} \mathcal{B} &= \mu \mathcal{H} & \mathcal{I} &= \sigma \mathcal{E} \\ \mathcal{D} &= \epsilon \mathcal{E} & \mathcal{I} &= \rho \mathcal{I} \end{aligned}$$

where μ , ϵ , σ , and ρ may vary with temperature but, for linear circuit elements, are invariant with respect to potential differences across, or current through, the circuit element. In practice this restriction does not impose objectionable limitations, except that analyses based on the assumption of constant permeability imply that corrections or modifications of the linear circuit theory are required in dealing with circuit elements employing ferromagnetic materials. The principle of superposition (Sec. 87) applies only to systems composed of linear circuit elements.

Three conditions exist for formulating the circuit parameters of lumped circuit elements. In each case the property of storing or dissipating electrical energy is represented by a different circuit element. In conductors the electric fields are accompanied by pure conduction currents and the dissipation of electrical energy in the form of heat. These are characteristic properties of resistors. The electric fields in dielectrics are accompanied by displacement currents and the storage of electrical (potential) energy within the dielectric. This is a characteristic property of capacitors. Magnetic fields are associated with the conduction of currents and are accompanied by the storage of magnetic (kinetic) energy. This is a characteristic property of inductors.

The circuit properties of resistance, capacitance, and inductance as employed in electrical analysis are idealized concepts. As such they represent hypothetical and limiting cases that can be approached but never fully attained in the actual embodiment of mathematical ideas as physical pieces of equipment. No physical circuit element embodies purely and completely only that circuit property for which it is named. All embody more or less perfectly that circuit property for which they are named, but they also embody the remaining two circuit properties. In well-designed circuit elements the extent to which the element conforms to the mathematical ideal circuit property is a matter of design and application of the element. The extraneous circuit properties can usually be made so small as to be of minor importance, except where the frequency of variation of the voltage and current is high. In any case, the actual behavior of any physical circuit element can be represented to any desired degree of approximation or precision by suitable combinations of idealized circuit parameters.

The idealized circuit properties discussed in the following sections are those rigorously defined for the steady case. This conventional point of view fails to take in account certain matters, such as the depth of penetration and phase differences, which become important at very high frequencies and require modifications of the treatment given here for extension to high frequencies. For treatment of these more advanced topics, the reader is referred to other chapters of this handbook or to references giving more specialized treatments of this topic.¹

76. Resistance and Conductance. Resistance is the (scalar) property of an electric circuit, or of any body that may be used as a part of an electric circuit, which determines for a given current the rate at which electric energy is converted into heat or radiant energy and which has a value such that the product of the resistance and the square of the current gives the rate of energy conversion. In the general case, resist-

¹ These topics are treated in such works as G. P. Harnwell, "Principles of Electricity and Electromagnetism," McGraw-Hill, 1938; A. Hund, "Phenomena in High-frequency Systems," McGraw-Hill, 1936; L. Page, "Principles of Electricity," Van Nostrand, 1931; S. Ramo and J. R. Whinnery, "Fields and Waves in Modern Radio," Wiley, 1944.

ance is a function of the current, but the term is most commonly used in connection with circuits in which the resistance is independent of the current.

Conductance is the property of an electric circuit, or of a body that may be used as a part of an electric circuit, which determines for a given emf in the circuit or for a given potential difference between the terminals of a part of a circuit, the rate at which electric energy is converted into heat or radiant energy, and which has a value such that the product of the conductance and the square of the emf, or potential difference, gives the rate of energy conversion. In the general case, conductance is a function of the potential difference, but the term is most often used in connection with circuits in which the conductance is independent of the potential difference.

If an electric field is applied to the ends of a conductor of uniform cross section made of homogeneous material, a potential difference will exist across the conductor and a current will flow through it, the charges (electrons) flowing along the streamline of the electric field ($\theta = 0$), since they are assumed not to flow through the dielectric that surrounds the conductor. For the steady case, the current flow will be of uniform current density over each part of the cross-sectional area,¹ so that if s is the area normal to the direction of the conductor, $s = I/s$. If ρ is the resistivity of the homogeneous material of which the conductor is composed so that $\mathcal{E} = \rho s$, the potential difference between points A and B of the filamentary conductor, may be expressed as

$$E_{AB} = \int_B^A \mathcal{E} \cos \theta \, dl = \int_B^A \rho s \cos \theta \, dl = \int_B^A \rho \frac{I}{s} \cos \theta \, dl$$

$$= \int_B^A \rho \frac{I}{s} \, dl = \rho \frac{l_{AB} I}{s} = RI$$

where $\theta = 0$, and $R = \rho l/s$ is the resistance of the conductor of area s between the points A and B . From a physical point of view, the resistance of a conductor may be regarded as the opposition to the flow of the conduction current through a conductor of specified cross-sectional area, specified resistivity, and specified length. The resistivity, on the other hand, is an inherent property of the material of which the resistor is made and is in no way related to the dimensions of the conductor or resistor.

The resistance of a conductor of uniform cross section and made of homogeneous material may be expressed in terms of the electric field intensity and the current density, or in terms of their integrated effects, voltage, and current, or, alternatively, in terms of the conductor dimensions and resistivity of the material of which the resistor is made. From the foregoing results we have the relation,

$$R = \frac{E}{I} = \frac{\int \mathcal{E} \cos (\mathcal{E}, l) \, dl}{\int \mathcal{E} \cos (s, n) \, ds} = \frac{\int \mathcal{E} \cdot dl}{\int \sigma \mathcal{E} \cdot nds} = \frac{\mathcal{E} \int dl}{\sigma \mathcal{E} \int ds} = \frac{l}{\sigma s} = \rho \frac{l}{s}$$

where $\int \mathcal{E} \cdot dl$ and $\int s \cdot ds$ are the field equations for the steady-state case electric field intensity and current density, respectively. Although the resistance has been specified in terms of the rather special case of a conductor of uniform cross section composed of homogeneous material, this procedure is largely a matter of convenience; the ratio of the two field equations specifies the resistance for any material of whatever physical form the resistor may take.

In the derivation of the constant resistance (above), it was assumed that the material was homogeneous and, therefore, that the resistivity and conductivity were constant. For many materials, these characteristic properties of material are not constant but depend upon the current through, or the potential difference across, the conductor. For such cases, E and I are not linearly related, since the resistance has a unique value for different magnitudes of voltage across or current through it. Since the resistance is no longer constant, it is customary to define the nonlinear resistance as the limiting ratio of the voltage increment to the resulting current increment, as the latter approaches zero. The variational or differential resistance thus defined is

¹ This condition is not fulfilled for high frequencies.

usually designated by the letter r and is given mathematically by the relation,

$$r = \lim_{\Delta I \rightarrow 0} \frac{\Delta E}{\Delta I}$$

where the voltage increment across the resistance is ΔE and the corresponding current increment is ΔI .

The reciprocals of the ohmic resistance, R , and of the variational resistance, r , are sometimes more useful electrical quantities than R and r for circuit calculations. These reciprocal quantities are called the conductance (designated by $G = 1/R$) and the "variational conductance" ($g = 1/r$) and are defined by the equations

$$G = \frac{I}{E} = \frac{\iint \mathcal{E} \cos(\mathcal{E}, n) ds}{\int \mathcal{E} \cos(\mathcal{E}, l) dl} = \frac{\mathcal{E} \iint ds}{\rho \mathcal{E} \int dl} = \frac{s}{\rho l} = \frac{1}{R}$$

for linear conductors, or

$$g = \lim_{\Delta E \rightarrow 0} \frac{\Delta I}{\Delta E}$$

for nonlinear conductors.

Resistance is measured in volts per ampere or in ohms in the mks system, and conductance is measured in reciprocal ohms or in mhos.

The linear resistance and linear conductance express an electrical characteristic of those materials or devices obeying Ohm's law. This includes most good conductors encountered in radio practice, particularly those of metals. It does not include conductors composed of carbon, nonhomogeneous materials, or those operating over an appreciable temperature range. Ohm's law does not usually apply for semi-conductors. The variational resistance is commonly employed in such nonlinear devices as electron tubes, electric arcs, copper oxide, selenium, or other rectifiers, thyrite nonlinear conductors, and temperature-sensitive conductors. The resistivity of materials commonly employed in the construction of resistors for radio purposes is given in Table 7, Sec. 64.

77. Capacitance and Elastance. Capacitance is that property of a system of conductors and dielectrics which permits the storage of electricity when a potential difference exists between the conductors. The magnitude of the capacitance is expressed as the ratio of the quantity of electricity on the conducting elements to the potential difference between them. If Q is the charge stored and E is the difference of potential between the conductors, the capacitance is defined to be

$$C = \frac{Q}{E}$$

If there are two conductors in a homogeneous, isotropic medium between which an electric field intensity, \mathcal{E} , produces a difference of potential $E = \int \mathcal{E} \cos \theta dl$ and if the electric flux between the two conductors is $\psi = \iint \mathcal{D} \cos \theta' ds$ so that the charge on the conductor is $Q = K'\psi$, the capacitance is defined to be

$$C = \frac{Q}{E} = \frac{K'\psi}{E} = \frac{K' \iint_s \mathcal{D} \cos \theta' ds}{\int \mathcal{E} \cos \theta dl} = \frac{K'\epsilon \mathcal{E} \iint_s ds}{\mathcal{E} \int dl} = K' \frac{\epsilon s}{l}$$

where ϵ is the permittivity of the dielectric material, s is the surface area of the conducting plates bounding the dielectric, and l is the distance between the conducting surfaces. It is interesting to observe that this expression shows the physical dimensional factors upon which the capacitance is based. The permittivity and breakdown voltage of some representative dielectrics are given in Table 10.

The expression above gives the capacitance as defined for static conditions since it has been derived on the assumptions that \mathcal{D} and \mathcal{E} were time-invariant. Exactly the same expressions are obtained if \mathcal{D} and \mathcal{E} are functions of time, but in this case the equation obtained for capacitance may now be expressed for the dynamic cases in terms

of varying values of potential difference and displacement current. For the dynamic case the expression for capacitance becomes

$$C = \frac{[d(K'\psi)/dt]}{dE/dt} = \frac{K'(d\psi/dt)}{dE/dt} = \frac{K' \int \int_s (d\mathcal{D}/dt) ds}{dE/dt} = \frac{I_d}{dE/dt} = I_d \frac{dt}{dE}$$

from which $I_d = C(dE/dt)$; hence, by integrating, the voltage-current relation for any wave form becomes

$$E = \int \frac{I_d}{C} dt$$

For electric circuits, the displacement current in capacitors is the same as the conduction current in the leads to the capacitor so that $I_e = I_d$; therefore, we may deal only with the conduction current.

Table 10. Relative Permittivity and Breakdown Voltage of Typical Dielectrics Used in Radio Engineering

Material	Relative permittivity	Breakdown voltage, kv/cm
Ebonite.....	2.80	300-1,100
Glass.....	5.4-9.9	300-1,500
Gutta-percha.....	3.3-4.9	80-200
Mica.....	5.6-5.8	1,500-2,000
Paraffin.....	2.20	350
Porcelain.....	5.73	
Quartz, fused.....	3.9	
Rubber.....	150-500
Sulfur (amorphous).....	3.9	
Wood.....	4.2-7.8	

It is easy to show that the static and dynamic equations for the capacitance, as derived above, are identical. From the static definition, $Q = CE$, and since $Q = \int I dt$ it follows that $CE = \int I dt$. Hence, by differentiating, $I = C(dE/dt)$ which is the kinetic definition given above.

The capacitance, as defined above, is the total capacitance, since the total electric flux originating on one surface or conducting plate is assumed to flow to the opposite conducting plate or surface. If there are a number of conductors separated by a dielectric and so arranged that the flux from conductor A flows to other conductors, B, C, \dots, N , then the total capacitance of conductor A with respect to all other conductors is given by the expression,

$$C_A = \frac{K' \int \int_s \mathcal{D} \cdot n ds}{\int_B^A \epsilon \cdot dl} + \frac{K' \int \int_s \mathcal{D} \cdot n ds}{\int_C^A \epsilon \cdot dl} + \dots + \frac{K' \int \int_s \mathcal{D} \cdot n ds}{\int_N^A \epsilon \cdot dl} = C_{AB} + C_{AC} + \dots + C_{AN}$$

In devices intended to exhibit primarily the property of capacitance (capacitors or condensers), the capacitance is the total capacitance between two plates or other conductors separated by air, oil, mica, or other suitable gaseous, liquid, or solid dielectric. Partial capacitances are encountered when more than two conductors are separated by a dielectric, as when circuits are coupled through capacitors. It is also encountered in the capacitance between the electrodes of a multielement electron tube, and sometimes these interelectrode capacitances are responsible for undesired behavior of the tube, even though the capacitances are fairly small.

Sometimes it is more convenient to employ another circuit concept, the *elastance*, which is the reciprocal of the capacitance. The elastance, represented by the letter S ,

may be either the total or the partial elastance, corresponding to the total or the partial capacitance. The total elastance is defined by the electrostatic relation,

$$S = \frac{1}{C} = \frac{E}{Q} = \frac{K''E}{\psi} = \frac{K'' \int \epsilon \cos \theta dl}{\iint_s \mathfrak{D} \cos \theta ds} = \frac{K'' \int \epsilon dl}{\iint \mathfrak{D} \cdot n ds} = \frac{K'' \int \epsilon dl}{\epsilon \iiint ds} = \frac{K'' l}{\epsilon s}$$

or, by the kinetic or dynamic relation,

$$S = \frac{1}{C} = \frac{E}{\int I dt}$$

Capacitance is measured in coulombs per volt to which the name "farad" has been given. The farad is a large value of capacitance for engineering purposes, so that subdivisions of it are more commonly employed. A capacitor has a capacitance of 1 farad when a charge of 1 coulomb is sustained by a difference of potential of 1 volt between the conductors. The microfarad (10^{-6} farad, abbreviated μf) and the micromicrofarad (10^{-12} farad, abbreviated $\mu\mu\text{f}$) are units of capacitance commonly used in radio engineering. Sometimes the abbreviations mf and mmf are used for μf and $\mu\mu\text{f}$, but such practice is not in conformity with the usually accepted abbreviations for metric prefixes (see Sec. 111).

The unit of elastance is the reciprocal farad or *daraf*, which is the elastance represented by 1 volt between two conductors whose charge is 1 coulomb.

78. Inductance. Inductance is the (scalar) property of an electric circuit, or of two neighboring circuits, which determines the emf induced in one of the circuits by a change of current in either one.

Self-inductance is the property of an electric circuit that determines, for a given rate of change of current in the circuit, the emf induced in the same circuit, the induced voltage, E , and the self-inductance, L , being related to the current flowing in the same circuit by the relation

$$E_1 = -L \frac{dI_1}{dt}$$

where L is the coefficient of self-inductance.

Mutual inductance is the common property of two associated electric circuits that determines, for a given rate of change of current in one of the circuits, the emf induced in the other. Thus, if the mutual inductance, M , is positive,

$$E_1 = -M \frac{dI_2}{dt}$$

$$E_2 = -M \frac{dI_1}{dt}$$

where E_1 and I_1 are the voltage induced and the current, respectively, in circuit 1, and E_2 and I_2 are the voltage induced and the current in circuit 2, and M is the coefficient of mutual inductance.

Inductance may be interpreted in several ways. We may, for example, interpret it solely in terms of the mathematical relations given above, we may aim to provide some physical pictorialization of the essential circuit phenomena taking place, or we may build up a concept of it on the basis of field notions.

79. Self-inductance. The relation between induced voltage, current, and self-inductance may be obtained directly from the definition of self-inductance and induced emf. The inductance is defined as the ratio of the flux linkages to the current, or $L = \Lambda/I = N\phi/I$, whereas the induced emf is equal to the time rate of decrease of flux linkages, or $E = -(d\Lambda/dt) = -[d(N\phi)/dt]$. From the first of these relations we have $N\phi = LI$ and, combining this with the second relation, there results,

$$E = -\frac{d(N\phi)}{dt} = -\frac{d(LI)}{dt} = -\left(L \frac{dI}{dt} + I \frac{dL}{dt}\right)$$

If the inductance does not vary with time, the familiar result is obtained

$$E = -L \frac{dI}{dt}$$

The general expression for induced emf may be used to derive the expression for the inductance in a way that gives a somewhat better picture of the physical action taking place. In a multiturn coil of N turns in a magnetic field whose flux is ϕ webers not all the turns cut all the flux lines even when the turns are close together. If K is the fraction of the flux linkages out of the total maximum number that is possible, the number of actual flux linkages is $KN\phi$. For most coils in radio work (except those in which mutual coupling between portions of the coil is adjustable or variable, as in variometers, for example) K and N are constant. For this special but very common case, the induced voltage is

$$E = -KN \frac{d\phi}{dt}$$

The magnetic flux may be produced by current flowing through the windings of the coil, and in this case the magnetic flux produced is given by $\phi = NI$. Hence the voltage induced in the coil by current flowing through it is given by the relation

$$E = -KN^2 \frac{dI}{dt} = -L \frac{dI}{dt}$$

from which $L = KN^2$. From this point of view the self-inductance is an electrical quantity whose magnitude depends upon the square of the number of turns of the coil and also upon the fraction of the total flux linkages that actually exist for the coil in question.

The self-inductance can also be derived from field relations. Provided it is not in a region of ferromagnetic materials, *i.e.*, so long as μ is constant, the inductance is the constant ratio of the magnetic flux linkages to the current flowing. Hence, from the field relations for the steady case, the self-inductance may be defined by the relation

$$L = \frac{\phi}{I} = \frac{\int \int \mathfrak{B} \cos \theta' ds}{K''' \int \mathfrak{H} \cos \theta dl}$$

For a solenoid of N turns (the solenoid being infinitely long so that end effects may be neglected) the number of flux linkages is $KN\phi$, and the self-inductance of the multiturn coil is then

$$L = \frac{KN \int \int \mathfrak{B} \cos \theta' ds}{K''' \int \mathfrak{H} \cos \theta dl}$$

For a coil whose length, l , is very much greater than its diameter, the value of \mathfrak{B} at the central axis is $\mathfrak{B} = \mu NI/l$, so that for such a coil the inductance is given by

$$L = \frac{KN \int \int \frac{\mu NI}{l} \cdot ds}{K''' \int \mathfrak{B} dl} = \frac{KN^2 \mu I s}{K''' I l} = \frac{KN^2 \mu s}{K''' l}$$

In the form here given, this formula is not useful for computing the self-inductance of solenoids unless the individual factors can be evaluated. Nevertheless, this result is useful for illustrating the physical factors and dimensions upon which the self-inductance of a solenoid depends.

Self-inductance is measured in henrys or convenient decimal submultiples of the henry, such as the millihenry or the microhenry.

A few authors sometimes make use of the reciprocal inductance, designated by the symbol Γ . The reciprocal inductance is, of course, defined to be $\Gamma = I/\phi$ and is measured in reciprocal henrys or *yrnehs*.

80. Mutual Inductance. The determination of mutual inductance requires consideration of two adjacent loops. If the current flows through one coil or loop, it establishes a magnetic field around its own turns, but it also establishes a magnetic field around the adjacent loop or coil in which it induces an emf. Conversely, the current through the second coil sets up a magnetic field which induces an emf in the first coil as well as in its own turns. Let I_1 and I_2 be the current in coils 1 and 2, respectively. Let the magnetic flux linking circuit 2 as a result of current in circuit 1 be ϕ_{21} , and let ϕ_{12} represent the magnetic flux linking circuit 1 as a result of current in coil 2. Then, if K is the fraction of the number of lines of magnetic flux set up by circuit 1 which cuts circuit 2 (or conversely) and if N_1 and N_2 are the number of turns of coils 1 and 2, respectively, the interlinking flux is given by either of the relations

$$\phi_{21} = KN_2I_1 \quad \text{or} \quad \phi_{12} = KN_1I_2$$

The coefficient of mutual inductance, M , is defined as the number of flux linkages of one circuit per unit current flowing in the other. Hence, mathematically, the coefficient of mutual inductance is given by

$$M = \frac{N_1\phi_{21}}{I_1} = \frac{N_2\phi_{12}}{I_2} = KN_1N_2$$

The magnitude of the mutual inductance between two circuits depends upon the number of turns in each of the two circuits as well as upon the shapes of the coils, their separation (linear as well as angular), and other factors having a bearing on the flux linkages.

Mutual inductance may be derived from the basic field concepts. The calculations are long and not particularly easy to carry out for any particular physical configuration of conductors. A treatment of this subject may be found in Harnwell.¹

The units in which mutual inductance is measured are the same as those for self-inductance: the henry and its decimal submultiples. But, whereas self-inductance is a positive quantity, mutual inductance may be either positive or negative.

81. Power Dissipated in a Resistor. If a potential difference, E , is maintained across a semiconductor (resistor), electric charges in the resistor will be set in motion. If a charge, dQ , is set in motion by the potential difference, E , the energy expended in the process is $dU = E dQ$, and if this element of charge traverses the resistor in time dt , the power expended in the process is

$$P = \frac{dU}{dt} = E \frac{dQ}{dt} = EI$$

For linear resistors that obey Ohm's law, $E = IR$; hence, the power dissipated in such a resistor is given by

$$P = EI = I^2R = \frac{E^2}{R}$$

The power dissipated will be measured in watts if I is the current in amperes, E is the potential drop in volts, and R is the resistance in ohms.

82. Energy Stored in Dielectric of a Capacitor. Suppose that at any instant of time the potential difference between the conducting elements of a capacitor has built up to the value E as a result of transferring an element of charge, dQ , to the plates or conductors. The amount of energy required for the process is $dU = E dQ$. Since, by definition of capacitance, $C = Q/E$, it follows that $Q = CE$ or $dQ = C dE$. Hence the energy increment is $dU = E dQ = CE dE$. Upon integrating, the total energy is found to be

$$U = \int dU = \int_0^E CE dE = 0.5CE^2 = 0.5 \frac{Q^2}{C}$$

The energy, in mks units, is in joules, E is in volts, Q in coulombs, and C in farads.

¹ Harnwell, G. P., "Principles of Electricity and Electromagnetism," p. 297, McGraw-Hill, 1938.

83. Energy Stored in Magnetic Field of Inductor. From the relations given in Sec. 78, the drop in potential difference across an inductor is given by the relation, $E = L(dI/dt)$, from which is obtained the expression, $E dt = L dI$. The energy expended in time dt is given by the relation $dU = EI dt = LI dI$. Hence, the total energy in the magnetic field of an inductor is the integral of dU , or

$$U = \int dU = \int_0^I LI dI = 0.5LI^2 = 0.5L \left(\frac{dQ}{dt} \right)^2$$

In mks units, the energy is in joules, I is in amperes, and L is in henrys, and Q is in coulombs.

The voltage-current relations for linear circuit elements such as have been treated here are summarized in Table 11, Sec. 96.

84. Circuit Laws and Theorems. The behavior of electrical circuits can be predicted on the basis of two relations originally stated by Kirchhoff and applicable under all conditions, and a relation, originally stated by Ohm which, although not universally applicable (at least in its simplest form), is true for many instances and is highly important and useful. In addition to these Kirchhoff and Ohm relations, a number of theorems find extensive usefulness in simplifying the calculations of circuit behavior. The most important of these are the principle of superposition, the compensation (substitution) theorem, the reciprocity theorem, the maximum power theorem, Thévenin's theorem, and Norton's theorem. These laws and theorems are treated in the following sections.

85. Kirchhoff's Laws. The two circuit laws of Kirchhoff may be stated as follows:

1. The algebraic sum of the currents flowing toward any point in an electric circuit is zero.
2. The algebraic sum of the product of the current and resistance in each of the conductors in any closed path in an electric circuit is equal to the algebraic sum of the emfs in that path.

As stated here, Kirchhoff's laws apply to the instantaneous values of the current and potential differences, but they may be extended to the effective values of sinusoidal currents and voltages by replacing "algebraic sum" by "vector sum" and replacing the term "resistance" by the term "impedance."

Each of the two Kirchhoff laws gives rise to a separate method of analyzing circuit behavior: the node and the loop method. In the node or junction method, the flow of current of all conductors meeting at a common point must be zero. Hence, if the magnitude and direction of all currents but one are known (or can be assumed), the remaining current can be determined by the application of the Kirchhoff law for current at a junction. In the loop method, the sum of all potential differences must be zero, or the sum of the rises of potential difference must equal the sum of the drops in potential difference.

86. Ohm's Law. Ohm's law states that the current in an electric circuit is directly proportional to the emf in the circuit. This law does not apply to all circuits but to metallic circuits and to many circuits containing electrolytic resistance. Ohm's law was first enunciated for a circuit in which there is constant emf and an unvarying current. It is applicable to circuits in which varying currents flow if account is taken of the induced emf resulting from the self-inductance of the circuit and of the distribution of current in the cross section of the circuit. Ohm's law applies only to linear circuits.

87. Principle of Linear Superposition. According to the principle of superposition, if several causes act simultaneously in a linear system, the resultant effect of all of them is equal to the sum of the individual effects of each cause acting by itself. As applied to electric circuits, the principle may be stated more explicitly. In a linear circuit, each emf acts independently of all other emfs in producing currents in the circuit.

88. Compensation Theorem. The compensation (or substitution) theorem states the conditions under which an emf may be replaced, in an electric circuit, by a poten-

tial difference across a resistor, so far as the circuit currents are concerned. Thus, provided the emf equals the potential drop across the resistance and is oppositely directed, a resistance element in a circuit can be replaced by an emf, or vice versa.

89. Reciprocity Theorem. In any branch, *A*, of an electric circuit, the current produced by an emf acting in another branch, *B*, is equal to the current in branch *B* when its emf is transferred to branch *A*. This is the equivalent of saying that a resistanceless source of emf may be interchanged with a resistanceless current-measuring instrument without altering the reading or indication of the instrument.

90. Maximum Power Transfer Theorem. The maximum amount of electric power is delivered from a source to a load connected to it when the resistance of the load is equal to the internal resistance of the power source. If the source and load contain reactance as well as resistance, maximum power is transferred from the source to the load when the load resistance is equal to the internal resistance of the source and when the reactance of the load is equal in magnitude but opposite in sign to that of the source. Thus, maximum power is delivered from a source to a load when the load impedance is the conjugate of that of the source.

91. Thévenin's Theorem. If two terminals are available for any linear network, then, so far as external calculations of the system are concerned, they may be regarded

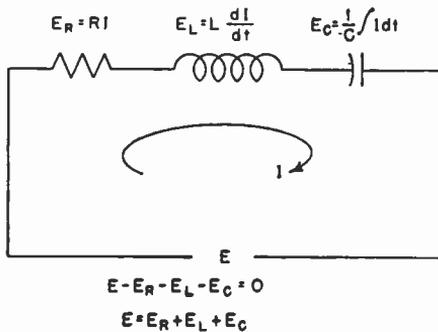


FIG. 26. Series circuit containing *R*, *L*, and *C*.

as the terminals of a simple series circuit having a resistance equal to the resistance measured between the terminals, and an emf equal to the difference of potential between the two terminals.

92. Norton's Theorem. In a linear circuit, any generator of electric power may be considered equivalent, at specified frequency, to a current generator whose current is equal to the short-circuit current in shunt with an admittance whose magnitude is equal to that which is measured across the generator when the generator is inactive and there is no load connected to it.

93. Electric Networks. An electric network is a system of interconnected circuit elements and sources of emf. The points at which two or more circuit elements are connected are called "nodes" or "junctions." A continuous path (or several conducting paths) through which current may flow is called a "loop."

94. Circuit Connections. Fundamentally, two types of circuits are possible, according to the way in which the circuit elements are joined:

1. In series-connected circuits, the circuit elements are connected, one after the other, in such a way that the same current flows, successively, through each element, and the applied emf is equal to the sum of the drops in potential differences across all the successively connected circuit elements.

2. In parallel- or shunt-connected circuits the circuit elements are connected side by side in such a way that the same emf is applied simultaneously to each and every element, the total current flowing through the circuit being the sum of the currents flowing through each circuit element individually.

Examples of the most general two-terminal series-connected circuit and the most general two-terminal shunt-connected circuit containing, in each case, one of the three common circuit elements are shown in Figs. 26 and 27, respectively.

Combinations of series-connected circuit elements and shunt-connected circuit elements are possible to give rise to series-shunt or shunt-series circuits of any degree of complexity, but in any case only these two methods of connection are possible for two-terminal circuit elements.

95. Resultant of Series- and Shunt-connected Circuit Elements. Circuit elements connected either in series or in parallel, provided they are all of the same kind, may be replaced by an equivalent circuit element whose circuit parameter is the resultant of that of the individually connected circuit elements. So long as there is no interaction between the fields of two or more circuit elements, *i.e.*, so long as no coupling exists between two or more circuit elements, the resultant circuit parameter of several like circuit elements connected together is simply expressed.

In the case of series-connected circuit elements, the resultant of several like circuit elements may be determined from the loop equations of Kirchhoff's laws. If subscripts 1, 2, . . . , *n* designate the component circuit elements connected in series without coupling or field interaction, the resultant circuit parameter, indicated by the subscript *T*, is given by the following relations:

For resistors,

$$R_T = R_1 + R_2 + R_3 + \dots + R_n = \sum_{n=1}^n R_n$$

For inductors having no mutual coupling,

$$L_T = L_1 + L_2 + L_3 + \dots + L_n = \sum_{n=1}^n L_n$$

For elastors (reciprocal capacitors) having no coupling,

$$S_T = S_1 + S_2 + S_3 + \dots + S_n = \sum_{n=1}^n S_n$$

In the case of parallel-connected circuit elements, the resultant of several like circuit elements may be determined from the node or junction equation of Kirchhoff's laws. If the subscript notation is the same as that used above for series circuits, the resultant circuit parameter, indicated by the subscript *T*, is given by the following relations:

For conductance elements (reciprocal resistors),

$$G_T = G_1 + G_2 + G_3 + \dots + G_n = \sum_{n=1}^n G_n$$

For reciprocal inductors having no mutual coupling,

$$\Gamma_T = \Gamma_1 + \Gamma_2 + \Gamma_3 + \dots + \Gamma_n = \sum_{n=1}^n \Gamma_n$$

For capacitors having no coupling,

$$C_T = C_1 + C_2 + C_3 + \dots + C_n = \sum_{n=1}^n C_n$$

In each case the resultant circuit element has properties like those of its constituent equivalent elements and is measured in the same electrical units of ohms, henrys, and

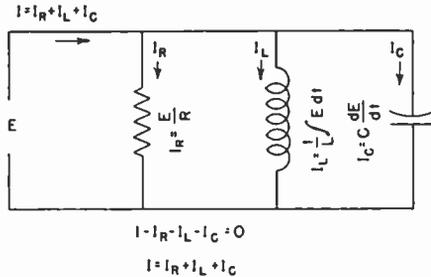


FIG. 27. Parallel circuit containing *R*, *L*, and *C*.

and *C*, indicated by the subscript *T*, is given by the following relations:

darafs for the series circuits, and in mhos, yrnehs, and farads for the shunt-connected circuit elements, respectively.

96. **Circuit Solutions.** The solution for an electric circuit consists in finding the relations between the voltages (rises as well as drops of potential difference) and currents as a function of time. This relationship is determined by means of Kirchhoff's laws and by writing the node or loop equations for the circuit in terms of the general voltage-current relations for circuit elements as given in Table 11, and including any sources of emf that may exist in the circuit.

Table 11. Voltage-Current and Energy Relations for Circuit Elements

Relation	Relation for the circuit property of		
	Resistance	Inductance	Capacitance
Voltage-current.....	$E = RI$	$E = L \frac{dI}{dt}$	$E = \frac{1}{C} \int I dt$
Current-voltage.....	$I = \frac{E}{R}$	$I = \frac{1}{L} \int E dt$	$I = C \frac{dE}{dt}$
Power dissipation.....	$P = I^2R = \frac{E^2}{R}$	$P = 0$	$P = 0$
Energy storage.....	$U = 0$	$U = \frac{1}{2}LI^2$	$U = \frac{1}{2}CE^2$

Since the general voltage-current relations for circuits composed of linear circuit elements are expressible through the use of differential, algebraic, or integral expressions, it is possible to differentiate these expressions with respect to time so that the circuit behavior is expressed in terms of differential equations that represent the dynamic equilibrium of the circuit.

The complete solution of the differential equation representing the dynamic equilibrium of the electric circuit consists of two parts: (1) a transient (represented or expressed by the particular integral solution of the differential equation) specifying the temporary redistribution of charges (hence, a temporary voltage and current component) resulting from a sudden change in dynamic conditions, and (2) a steady-state solution (expressed by the complementary function of the differential equation) which represents the dynamic equilibrium of the circuit when subjected to an applied emf.

The currents and voltages representing the transient conditions of a network usually decrease rapidly with time and soon become negligible. In the steady-state condition, the voltages and currents may vary cyclically, but their effective value is constant. In the case of a harmonically applied emf, the drops in potential difference across each of the circuit elements and the current through the linear circuit elements are also harmonic functions of time and of the same frequency (but not the same phase, in general) as that of the applied emf. It is not generally true for any wave form, however, that the same wave shape exists for each of the circuit elements in the circuit.

97. **Steady-state Solution.** The steady-state condition of an electric circuit composed of linear two-terminal elements of R , L , and C is that in which the effective value of the current flowing in the circuit as a result of an applied emf is constant, each cycle of the recurrent wave form being like that of every other cycle. In general, the wave forms of the voltages across the different kinds of circuit elements of a series-connected circuit are unlike those across other circuit elements and, except in the case of a resistor, are not like the wave form of the applied emf. Likewise in shunt-connected circuits, the wave form of current through each circuit element is, in general, different from that flowing through circuit elements of another kind. Except for phase displacements, the wave forms are alike in the special but important case of harmonic voltages and currents. This is a direct consequence of the fact that the derivative and integral of a harmonic function are also harmonic functions.

The determination of the steady-state solution of electric circuits composed of linear circuit elements will be illustrated by finding the current that flows in a series R, L, C circuit when a harmonic emf, $E = E_m e^{j\omega t}$ is impressed across the series circuit of Fig. 26. From Kirchhoff's voltage or loop law and the voltage-current relations for the three kinds of circuit elements in the system, it follows that the loop equation takes the form

$$E = E_m e^{j\omega t} = E_L + E_R + E_C = L \frac{dI}{dt} + RI + \frac{1}{C} \int I dt$$

where I is the current flowing in the circuit and whose value is to be determined. Since it is easier to manipulate differential equations rather than integro-differential equations, the loop equation above may be transformed into a differential equation by differentiating both sides with respect to t . The result is

$$j\omega E_m e^{j\omega t} = L \frac{d^2 I}{dt^2} + R \frac{dI}{dt} + \frac{1}{C} I$$

A solution of this equation is $I = I_m e^{j\omega t}$, from which $(dI/dt) = j\omega I_m e^{j\omega t}$ and $(d^2 I/dt^2) = -\omega^2 I_m e^{j\omega t}$. Substituting these values in the above equation and dividing both sides by $j\omega$, there is obtained

$$E_m e^{j\omega t} = \left(j\omega L + R + \frac{1}{j\omega C} \right) I_m e^{j\omega t} = Z I_m e^{j\omega t}$$

where $Z = [j\omega L + R + (1/j\omega C)] = R + j[\omega L - (1/\omega C)]$ is the impedance of the circuit. From this result it is evident that current and voltage both vary harmonically with time, as does the voltage drop across each of the circuit elements.

98. Transient Solution. The transient condition of an electric circuit composed of linear elements of R, L , and C is that which involves the temporary redistribution of charges as the result of a sudden alteration of the dynamic equilibrium of the electric circuit. Since this transient condition expresses a readjustment of the circuit condition, the solution does not involve any continuously externally applied emf or current. It does, however, involve the initial conditions of the circuit at the time when the circuit equilibrium is disturbed. The wave forms of voltage and current depend not only upon what elements are present in the circuit but also upon the relative magnitudes of the parameters of the circuit elements in the system.

The transient solution of electric circuits will be illustrated by finding the transient current that flows in a series circuit of linear R, L , and C circuit elements, when the quiescent equilibrium of the system is suddenly disturbed by closing the circuit with a battery of time-invariant emf, E , in the circuit. Physically, the circuit represents the charge on a capacitor in a circuit also containing inductance and resistance, under the assumption that the initial charge on the capacitor is zero.

At any instant of time, the Kirchhoff loop equation requires that

$$E = E_L + E_R + E_C = L \frac{dI}{dt} + RI + \frac{1}{C} \int I dt$$

To transform this to a differential equation, differentiate with respect to t to obtain

$$\frac{dE}{dt} = 0 = L \frac{d^2 I}{dt^2} + R \frac{dI}{dt} + \frac{1}{C} I$$

the general solution of which is

$$I = \left[\frac{E}{\left(R^2 - \frac{4L}{C} \right)^{1/2}} \right] \epsilon^{-(Rt/2L)} \left[\epsilon^{\sqrt{R^2 - \frac{4L}{C}} t} - \epsilon^{-\sqrt{R^2 - \frac{4L}{C}} t} \right]$$

From this solution, it is evident that the term with the negative exponent indicates that the transient current will approach zero for a sufficiently large value of time,

approaching zero current at infinite time. But the manner in which this zero current is approached depends upon the relative values of R , L , and C , and three different cases may exist for this circuit:

1. If $R^2 > 4L/C$, the exponential terms in the brackets are positive numbers; the current begins at zero, rises to a maximum value, and then gradually decays to zero slowly, without changing polarity.

2. If $R^2 = 4L/C$, the solution above becomes indeterminate, but by differentiating both numerator and denominator individually with respect to the factor that is the coefficient of t , the solution for the current is

$$I = \left(\frac{Et}{L} \right) e^{-(Rt/2L)}$$

For this case, often called the "critical" case, the current begins at zero, rises to a maximum value, and returns again to zero without changing polarity. This is the limiting condition of case 1 in which the rise and fall occur more slowly than for the present critical case.

3. If $R^2 < 4L/C$, the current takes the form

$$I = \left(\frac{2E}{\sqrt{\frac{4L}{C} - R^2}} \right) e^{-(Rt/2L)} \sin \omega t$$

where $\omega = \frac{1}{2L} \left(\frac{4L}{C} - R^2 \right)^{1/2}$. For this case it is evident from the equation above that the current oscillates with an angular velocity ω and that the amplitude steadily decreases with time until the magnitude of the transient oscillations becomes negligible, ultimately reaching zero at infinite time.

Transients are assuming increasing importance in radio engineering design, and the following list of references will serve as a suitable introduction to this topic.

References

- Berg, E. J.: "Heaviside's Operational Calculus," McGraw-Hill, 1936.
 Bush, V.: "Operational Circuit Analysis," Wiley, 1929.
 Carson, J. R.: "Electric Circuit Theory and the Operational Calculus," McGraw-Hill, 1926.
 ———, and O. J. Zobel: Transient Oscillations in Electric Wave Filters, *Bell System Tech. J.*, 2 (3), 1, July, 1923.
 Gardner, M. F., and J. L. Barnes: "Transients in Linear Systems," Wiley, 1942.
 Guillemin, E. A.: "Communication Networks," Vol. 1, Wiley, 1931.
 Kerchner, R. M., and G. F. Corcoran: "Alternating Current Circuits," Wiley, 1938.
 Kurtz, E. B., and G. F. Corcoran: "Introduction to Electric Transients," Wiley, 1935.
 Massachusetts Institute of Technology, Electrical Engineering Dept.: "Electric Circuits," Wiley, 1940.
 McIlwain, K., and J. G. Brainerd: "High Frequency Alternating Currents," Wiley, 1931.
 Pierce, G. W.: "Electric Oscillations and Electric Waves," McGraw-Hill, 1920.
 Russell, A.: "Alternating Currents," Cambridge, 1914.
 Skilling, H. H.: "Transient Electric Currents," McGraw-Hill, 1937.

99. Impedance, Resistance, and Reactance. The impedance of an electric circuit, or portion of an electric circuit, to a completely specified periodic potential difference and current is the ratio of the effective or rms value of the potential difference across the circuit to the effective value of the current flowing through it, there being no source of power in the portion under consideration. The impedance is designated by the letter symbol, Z , and if E_{rms} and I_{rms} are the effective values of the potential drop and the current of the circuit, the circuit impedance is given by

$$Z = \frac{E_{rms}}{I_{rms}}$$

The impedance is measured in ohms.

If the periodic emf applied to the circuit is composed of components of different frequency, if $E_0, E_1, E_2, \dots, E_n$ represent the effective values of voltages of frequencies $0, \omega, 2\omega, \dots, n\omega$, if $I_0, I_1, I_2, \dots, I_n$ represent the effective values of the currents of the same frequencies, respectively, and if the maximum values of the voltage are $E_0, E_{1m}, E_{2m}, \dots, E_{nm}$ and the maximum (amplitude) values of the currents

are $I_0, I_{1m}, I_{2m}, \dots, I_{nm}$, then the impedance of the circuit or portion of the circuit is given by the relation

$$Z = \frac{E_{rms}}{I_{rms}} = \frac{\sqrt{2E_0^2 + E_{1m}^2 + E_{2m}^2 + \dots}}{\sqrt{2I_0^2 + I_{1m}^2 + I_{2m}^2 + \dots}} = \frac{\sqrt{E_0^2 + E_1^2 + E_2^2 + \dots}}{\sqrt{I_0^2 + I_1^2 + I_2^2 + \dots}}$$

The impedance thus defined is a scalar quantity and gives merely the ratio of the magnitudes of the voltage and current.

The vector impedance of an electric circuit, or portion of a circuit, for simple harmonic voltage and current is the ratio of the corresponding complex harmonic potential difference to the corresponding complex current. In the example given in Sec. 97, the applied harmonic voltage is $E = E_m e^{j\omega t}$ and the resulting current is $I = I_m e^{j\omega t}$ and the vector impedance. *i.e.*, the impedance expressed as a vector in rectangular coordinates, is

$$Z = \frac{E_m e^{j\omega t}}{I_m e^{j\omega t}} = \left[R + \left(j\omega L + \frac{1}{j\omega C} \right) \right] = \left[R + j \left(\omega L - \frac{1}{\omega C} \right) \right] = R \pm jX$$

where R is the resistance, $j\omega L$ is the inductive reactance in ohms, $1/j\omega C = -j/\omega C$ is

the capacitive reactance in ohms, and $X = j \left(\omega L - \frac{1}{\omega C} \right)$ is the net reactance in ohms.

Depending upon whether the inductive or the capacitive reactance has the larger magnitude, the net reactance will be, respectively, positive or negative, but the resistance will always be positive for linear dissipative circuits.

In the above expression the impedance has been given in rectangular components, R and X , but the impedance may also be expressed in polar coordinates in terms of the magnitude of the impedance, $|Z|$, and the angle of phase shift, θ of the impedance.

Thus, in polar coordinates, the impedance for the series R, L, C circuit is

$$\begin{aligned} Z &= |Z|/\theta = \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C} \right)^2} / \tan^{-1} \frac{R}{\left(\omega L - \frac{1}{\omega C} \right)} \\ &= \sqrt{R^2 + X^2} / \tan^{-1} (R/X) \end{aligned}$$

where $|Z| = \sqrt{R^2 + X^2}$ and $\theta = \tan^{-1} (R/X)$. Conversely, the rectangular components of the impedance, in terms of the polar components, are

$$\begin{aligned} R &= |Z| \cos \theta \\ X &= |Z| \sin \theta \end{aligned}$$

Physically the impedance of a circuit is the opposition to the flow of a periodic current under steady-state conditions. It is composed of two parts: a resistive component usually regarded as constant (although the resistance is, strictly speaking, a function of frequency) and a reactive component whose magnitude usually varies appreciably with frequency. Over a sufficiently large frequency range the sign of the reactance also changes. The resistive component accounts for the absorption or dissipation of power. The reactive component results from properties of inductors and capacitors that involve the voltage-current rate of change relationships. For ideal circuit elements the reactance determines the frequency selective properties of the circuit.

The impedance is an electrical quantity whose use is most convenient in solving electrical circuits with circuit elements connected in series.

100. Admittance, Conductance, and Susceptance. The admittance of an electric circuit is the reciprocal of the impedance; it is, therefore, the ratio of the effective value of the current through a circuit or portion of a circuit, to the effective value of the potential difference across the circuit, for a completely specified periodic applied

voltage. The admittance is designated by the letter Y and is measured in reciprocal ohms or mhos or its multiples or submultiples.

If E_{rms} and I_{rms} are the effective values of the voltage and current, respectively, the admittance is given by

$$Y = \frac{I_{rms}}{E_{rms}}$$

For a nonharmonic periodic voltage, the admittance may be determined in a manner similar to that for determining the impedance of a circuit in which nonharmonic current flows.

In any electric circuit whose resistive and reactive components are R and X , respectively, the corresponding admittance is given by the relation

$$Y = \frac{1}{Z} = \frac{1}{R \pm jx} = \frac{1}{R \pm jx} \cdot \frac{R \mp jx}{R \mp jx} = \frac{R}{R^2 + X^2} \mp j \frac{X}{R^2 + X^2} = G \mp jB$$

where $G = R/(R^2 + X^2) = R/Z^2$ is the conductance of the circuit (measured in mhos) and $B = \mp X/(R^2 + X^2) = X/Z^2$ is the susceptance of the circuit measured in mhos.

The admittance may be expressed in polar components as well as in the rectangular components given above. In polar components,

$$Y = |Y| \angle \theta = \sqrt{G^2 + B^2} / \tan^{-1}(B/G)$$

where $|Y| = \sqrt{G^2 + B^2}$ and $\phi = \tan^{-1}(B/G)$. Conversely, the rectangular components of the admittance, in terms of polar components, are

$$\begin{aligned} G &= |Y| \cos \phi \\ B &= |Y| \sin \phi \end{aligned}$$

Since impedance and admittance are reciprocally related, the impedance may be expressed in terms of admittance components just as admittance was expressed, above, in terms of impedance components. Thus, in terms of the admittance, the impedance is

$$Z = \frac{1}{Y} = \frac{1}{G \pm jB} = \frac{1}{G \pm jB} \cdot \frac{G \mp jB}{G \mp jB} = \frac{G}{G^2 + B^2} \mp j \frac{B}{G^2 + B^2}$$

Physically, the admittance expresses the ease with which the current flows through a circuit. It is composed of two parts: a conductance component, which is usually regarded as constant and which enters into determinations of power, and a susceptive component, which usually varies appreciably with frequency in a manner inverse to that of the reactance of the circuit.

Admittance is a concept that is most useful in dealing with parallel circuits.

101. Resonance. Resonance exists between one coordinate of a system that is executing oscillations or vibrations and a periodic agency that maintains the oscillations or vibrations when a small amplitude of the periodic agency produces in the system a relatively large amplitude of the oscillation or vibration.

Amplitude resonance exists when the resonance is such that any change in the period of the periodic agency without changing its amplitude produces a decrease in the amplitude of the oscillation or vibration of the system.

Period or natural resonance exists when the resonance is such that the period of the applied agency is the same as the natural period of oscillation or vibration of the system.

Phase or velocity resonance exists when the resonance is such that the angular phase difference between the fundamental components of the oscillation or vibration and of the applied agency is 90 deg ($\pi/2$ radians). With phase resonance, the time derivative of the coordinate that is in resonance is a maximum.

Resonance may exist in an electric circuit possessing inductance, resistance, and capacitance between the quantity of electricity that oscillates and a periodically applied emf that sustains the oscillations. Any one of three kinds of resonance, *viz.*, amplitude resonance, period resonance, or phase resonance, may exist depending

upon the constants of the circuit. In electric circuits having a high degree of selectivity, there is little distinction between the three kinds of resonance.

In a series circuit composed of linear circuit elements of R , L , and C , resonance is obtained when the net reactance is 0 or when $\omega L = 1/\omega C$. Hence, the frequency of resonance is obtained from the relation, $\omega^2 = 1/(LC)$ so that

$$f_r = \frac{1}{2\pi \sqrt{LC}}$$

The frequency will be expressed in cycles per second if L is inductance in henrys and C is capacitance in farads.

References

- Albert, A. L.: "Electrical Fundamentals of Communication," McGraw-Hill, 1942.
 Berg, E. J.: "Electrical Engineering," McGraw-Hill, 1929.
 Carson, J. R.: "Electric Circuit Theory and the Operational Calculus," McGraw-Hill, 1926.
 Everitt, W. L.: "Communication Engineering," McGraw-Hill, 1937.
 Franklin, W. S., and F. E. Terman: "Transmission Line Theory," Franklin and Charles, 1926.
 Frazier, R. H.: "Elementary Electric-circuit Theory," McGraw-Hill, 1945.
 Glasgow, R. S.: "Principles of Radio Engineering," McGraw-Hill, 1936.
 Guillemain, E. A.: "Communication Networks," Vols. I and II, Wiley, 1931 and 1935.
 Hague, B.: "Alternating Current Bridge Methods," Pitman, 1943.
 Heaviside, O.: "Electrical Papers," Copley, Boston, 1925.
 Johnson, K. S.: "Transmission Circuits for Telephone Communication," Van Nostrand, 1925.
 Kennelly, A. E.: "Electric Lines and Nets," McGraw-Hill, 1928.
 Kerchner, R. M., and G. F. Corcoran: "Alternating Current Circuits," Wiley, 1943.
 Knight, A. R., and G. H. Fett: "Introduction to Circuit Analysis," Harper, 1943.
 Lawrence, R. R.: "Principles of Alternating Currents," McGraw-Hill, 1935.
 Malti, M. G.: "Electrical Circuit Analysis," Wiley, 1930.
 Massachusetts Institute of Technology, Electrical Engineering Dept.: "Electric Circuits," Wiley, 1940.
 McIlwain, K., and J. G. Brainerd: "High Frequency Alternating Currents," Wiley, 1931.
 Morecroft, J. H.: "Principles of Radio Communication," Wiley, 1927.
 Pierce, G. W.: "Electric Oscillations and Electric Waves," McGraw-Hill, 1920.
 Russell, A.: "Theory of Alternating Currents," Cambridge, 1914.
 Schwaiger, A.: "Theory of Dielectrics," Wiley, 1932.
 Shea, T. E.: "Transmission Networks and Wave Filters," Van Nostrand, 1929.
 Starr, A. T.: "Electric Circuits and Wave Filters," Pitman, 1934.
 Terman, F. E.: "Radio Engineering," McGraw-Hill, 1947.
 Ware, L. A., and M. R. Reed: "Communication Circuits," Wiley, 1944.
 Weinbach, M. P.: "Alternating Current Circuits," Macmillan, 1933.

MAGNETIC CIRCUITS

102. Electric and Magnetic Circuits. A magnetic circuit is a closed path of magnetic flux, the path having the direction of the magnetic induction at every point. On the other hand, an electric circuit is a closed path of electric current, the path having the direction of the electric current at every point. Although the equations for electric and magnetic circuits may take the same mathematical form, the analogy between them expresses no physical correspondence. Particularly is this true for magnetic induction, which does not represent a dissipation of power, whereas the flow of electric current does.

Corresponding to the relation between voltage, current, and resistance, as given by Ohm's law for an electric circuit, the magnetic circuit relation may be written

$$\mathfrak{F} = \phi \mathcal{R}$$

where \mathfrak{F} is the magnetomotive force in ampere-turns in the mks system, ϕ is the magnetic flux or induction in webers, and \mathcal{R} is the reluctance of the magnetic circuit in ampere-turns per weber. An unvarying magnetic flux is regarded as a static condition, whereas in the equivalent electric circuit, the steady flow of current is the corresponding uniform flow of electric charges. Reluctance is not analogous to resistance in the sense of accounting for energy dissipation. Another difference between electric and magnetic circuits is the degree to which electric or magnetic effects are confined to their respective circuits. Whereas excellent insulators for the flow of electric conduction current exist, there is no insulator for magnetic induction; therefore, the magnetic flux is not so completely confined to its magnetic circuit as the conduction current is confined to the electric circuit. From the practical point of view, the most

useful magnetic circuits are those in which the lines of magnetic force flow through ferromagnetic materials, and for such materials, the reluctance is not constant; in this respect, too, magnetic circuits differ from most electric circuits for which Ohm's law is obeyed.

103. Magnetic Permeability and Reluctivity. Permeability is the property of an isotropic medium which, under specified conditions, determines the relation between the magnitudes of the magnetizing force and the magnetic induction of the medium. Under the specified conditions, permeability is measured as the ratio of the magnetic induction to the magnetizing force, or

$$\mu = \frac{\mathfrak{B}}{\mathfrak{H}}$$

where, in mks units, \mathfrak{B} is the magnetic flux density or induction in webers per square meter, and \mathfrak{H} is the magnetic field intensity in ampere-turns per meter.

The relative permeability of most substances does not differ appreciably from that of vacuum, which is taken as unity. For practical purposes the relative permeability is constant, independent of the flux density. Iron, nickel, steel, cobalt, and magnetic alloys—the materials of practical importance in magnetic circuits—have relative permeabilities appreciably greater than unity, and the permeability varies considerably with the magnetizing force or the flux density, with the composition, and with the previous heat-treatment of the magnetic material.

For a uniform magnetic path, the reluctance may be determined from the field expressions for the magnetomotive force and magnetic flux,

$$\mathcal{R} = \frac{\mathfrak{F}}{\phi} = \frac{\int \mathfrak{H} \cos \theta \, dl}{\int \int_s \mathfrak{B} \cos \theta' \, ds} = \frac{\mathfrak{H} \int dl}{\mu \mathfrak{H} \int \int_s ds} = \frac{l}{\mu s} = \nu \frac{l}{s}$$

where \mathfrak{F} is the magnetomotive force in ampere-turns, ϕ is the magnetic flux in webers, \mathfrak{H} is the magnetic field intensity in ampere-turns per meter, θ is the angle between \mathfrak{H} and the normal to the surface s and for this case has the value $\theta = 0$, μ is the permeability in henrys per meter, and $\nu = 1/\mu$ is the reluctivity in meters per henry. For nonmagnetic substances (vacuum, air, and so on) the permeability and reluctivity are usually written μ_0 and ν_0 . The permeability and reluctivity of magnetic materials are not usually the absolute permeability and reluctivity, but the relative values of these quantities referred to the values for vacuum are those customarily tabulated.

In some magnetic circuits, varying as well as steady components of magnetic flux flow. The permeability offered to the varying component of the magnetizing force superimposed upon the steady component of the magnetizing force is called the "incremental permeability," the "apparent permeability," or sometimes the "a-c permeability." For such cases the magnetization cycle follows a small displaced hysteresis loop. The incremental permeability is measured by the slope of the straight line connecting the end points of this hysteresis loop. The incremental permeability decreases with increasing magnetization due to the steady component of magnetizing force and, up to the point at which saturation occurs, increases with the magnitude of the time-varying component of the magnetic flux density.

104. Reluctance and Permeance in Series and Parallel Magnetic Circuits. Permeance of a magnetic circuit is the reciprocal of the reluctance and expresses the ease with which magnetic flux may be established in a magnetic circuit, whereas reluctance is a measure of the opposition or difficulty with which magnetic flux is established in the magnetic circuit.

For magnetic circuits in series with one another, the total reluctance is the sum of the reluctance of the individual circuits, plus the reluctance of the joints where the magnetic circuits of high permeability are joined. If magnetic circuits are placed in parallel with one another, the permeance of each circuit is added to ascertain the total or net permeance of the combination of magnetic circuit branches in parallel.

References

Abraham, M., and R. Becker: "Classical Theory of Electricity and Magnetism," Blackie, 1932
 Bitter, F.: "Ferromagnetism," McGraw-Hill, 1937.
 Ewing, J. A.: "Magnetic Induction in Iron and Other Substances," Van Nostrand, 1900.
 Massachusetts Institute of Technology, Electrical Engineering Dept.: "Magnetic Circuits and Transformers," Wiley, 1943.
 Moullin, E. B.: "Principles of Electromagnetism," Oxford, 1932.
 Spooner, T.: "Properties and Testing of Magnetic Materials," McGraw-Hill, 1931.
 Stoner, E. C.: "Magnetism and Matter," London, 1934.
 Williams, S. R.: "Magnetic Phenomena," McGraw-Hill, 1931.

ELECTROMAGNETIC WAVES

106. Electromagnetic Waves. A wave is a disturbance that is propagated in a medium in such a way that, at any point in the medium, the displacement of the wave is a function of the position of the point. Any physical quantity that has the same relationship to some independent variable (usually time) that a propagated disturbance has at a particular instant, with respect to space, may be called a "wave." The acceleration of electric charges is capable of producing waves in the electromagnetic field, and it is these electromagnetic waves that make possible radio communication without the intervention of physical conductors between the transmitting and the receiving stations.

Maxwell's equations express the general relations between electrical and magnetic quantities, ϵ , \mathcal{D} , \mathcal{H} , and \mathcal{G} ; hence, the Maxwell equations may be expected to yield the means for deriving the equations of electromagnetic waves. This is indeed the case, but the variables are interconnected in such a manner that Maxwell's differential equations must be simplified (by separating the variables) before any solutions for electromagnetic waves can be ascertained. When the variables have been properly separated for the four Maxwell field equations through suitable mathematical manipulations, it is next required to find the relation or relations connecting time and the directional derivatives of ϵ , \mathcal{D} , \mathcal{H} , and \mathcal{G} , such that all four field equations are simultaneously satisfied and that the time-space relations will be in accordance with the definitions above for the propagation of electromagnetic waves. Since the first two (divergence) field equations do not involve time functions, the procedure is essentially that of finding a solution to the two circuital equations, both of which must be simultaneously satisfied.

In a medium that is perfectly insulating (zero conductivity) and free from charge density (free space satisfies these two conditions), the field equations to be solved,

$$\begin{aligned} \nabla \cdot \mathcal{E} &= 0 \\ \nabla \cdot \mathcal{H} &= 0 \\ \nabla \times \mathcal{E} &= -\dot{\mathcal{H}} \\ \nabla \times \mathcal{H} &= \dot{\mathcal{E}} \end{aligned}$$

Only electromagnetic waves that (when expressed in rectangular coordinates) satisfy these equations are possible; the equations above may, therefore, be used to determine the required form of the fields for wave propagation.

Solution of the above equations requires that each directional component of the electric or magnetic field intensity be of the form

$$\begin{matrix} \mathcal{E} \\ \diagdown \\ \diagup \\ \mathcal{H} \end{matrix} = f\left(t - \frac{l}{v}\right) + f\left(t + \frac{l}{v}\right)$$

where v is the velocity of propagation and is given by $v = 1/\sqrt{\mu\epsilon}$, l is used generically to designate a distance in any of the three orthogonal directions, and the functions, f , represent any single-valued functions. For propagation in free space, the propagation velocity has the special value of the velocity of light, approximately 3×10^8 m per sec obtained from the relation $v = c = 1/\sqrt{\mu_0\epsilon_0}$. The fact that the functions, f , are arbitrary functions of time makes it possible for the electromagnetic waves to be formed in such a way as to be capable of representing intelligence. The fundamental problem of radio communication is to make the time variations of the electromagnetic wave conform to the variations of the intelligence to be transmitted.

The energy stored in the electromagnetic field travels with the wave from the source of the disturbance outward into space, and this phenomenon is the mechanism by which the transmission of electric power, sufficient to carry on communication, occurs. An important point regarding the transmission of electromagnetic waves through space is that the energy is propagated through a dielectric; no conduction can occur in the (ideal) dielectric of free space. The mechanism of propagation is, briefly, that the variations of electric field intensity give rise to variations of magnetic field intensity which, in turn, give rise to variations of the electric field intensity once more, and so on. Once established, the original disturbance is sustained by the mutual interaction of electric and magnetic components of the electromagnetic field as the wave travels through space with the velocity of light. The amplitude of the variations of the electromagnetic field decreases with distance from the center of the disturbance as the wave expands to greater and greater size.

106. Energy in the Electromagnetic Field. Since, as noted above, the energy of the electromagnetic field radiates outward into space, it is desirable to have quantitative expressions for the field energy. Such expressions can be obtained from energy relations derived in Secs. 82 and 83 by converting the circuital relations into field relations.

From the relation giving the total energy stored in the dielectric of a capacitor, the total electric energy stored in an electromagnetic field of volume v is

$$\begin{aligned} U_{\mathcal{E}} &= 0.5CE^2 = 0.5QE = 0.5K'' \left(\int \int_v \mathfrak{D} \cdot n \, ds \right) \cdot \left(\int \mathcal{E} \cdot dl \right) \\ &= 0.5K'' \left(\int \int_s \epsilon \mathcal{E} \cdot n \, ds \right) \cdot \left(\int \mathcal{E} \cdot dl \right) = 0.5K'' \epsilon \mathcal{E}^2 \int \int \int dv = 0.5K'' \epsilon \mathcal{E}^2 v \end{aligned}$$

so that the electric energy per unit volume of field is

$$u_{\mathcal{E}} = 0.5K'' \epsilon \mathcal{E}^2$$

In a similar way, by making use of the energy in the magnetic field surrounding a coil, the total magnetic energy stored in an electromagnetic field of volume v is

$$\begin{aligned} U_{\mathcal{H}} &= 0.5LI^2 = 0.5 \left(\frac{K'' \int \int_s \mathfrak{D} \cdot n \, ds}{\int \mathcal{H} \cdot dl} \right) \cdot \left(\int \mathcal{H} \cdot dl \right)^2 \\ &= 0.5K'' \left(\int \int_s \mu \mathcal{H} \cdot n \, ds \right) \cdot \left(\int \mathcal{H} \cdot dl \right) = 0.5K'' \mu \mathcal{H}^2 \int \int \int dv = 0.5K'' \mu \mathcal{H}^2 v \end{aligned}$$

so that the magnetic energy per unit volume is

$$u_{\mathcal{H}} = 0.5K''\mu\mathcal{H}^2$$

The total energy, per unit volume, is the sum of the electric and magnetic energy, or

$$u = u_{\mathcal{E}} + u_{\mathcal{H}} = K''(0.5\epsilon\mathcal{E}^2 + 0.5\mu\mathcal{H}^2)$$

107. Poynting's Vector. Poynting's vector has importance in the determination of power flow throughout space; it is, therefore, frequently encountered in radiation phenomena. Usually designated by the letter S , it represents the energy per second flowing per unit area perpendicular to the direction of propagation.

At a point in space, the total flow of power (or energy per second) is equal to the product of the electric field intensity and the current density (representing a dissipation of power) plus the time rates of change of the magnetic field and the electric field. Mathematically, in terms of Poynting's vector, S ,

$$-\text{div } S = \mathcal{J}\mathcal{E} + \frac{\partial u_{\mathcal{E}}}{\partial t} + \frac{\partial u_{\mathcal{H}}}{\partial t}$$

This equation defines conditions at a point in the electromagnetic field, but it is possible to define the same conditions for a volume, dv . Thus, by expressing \mathcal{J} , \mathcal{E} , and \mathcal{H} in terms of their orthogonal components, we obtain the integral expression for rectangular coordinates,

$$\begin{aligned} & - \iiint_v \left(\frac{\partial s_x}{\partial x} + \frac{\partial s_y}{\partial y} + \frac{\partial s_z}{\partial z} \right) dv \\ & = \iiint_v (s_x \mathcal{E}_x + s_y \mathcal{E}_y + s_z \mathcal{E}_z) dv \\ & \quad + \iiint_v \frac{\partial}{\partial t} \left(\frac{\epsilon \mathcal{E}_x^2}{2} + \frac{\epsilon \mathcal{E}_y^2}{2} + \frac{\epsilon \mathcal{E}_z^2}{2} \right) dv \\ & \quad + \iiint_v \frac{\partial}{\partial t} \left(\frac{\mu \mathcal{H}_x^2}{2} + \frac{\mu \mathcal{H}_y^2}{2} + \frac{\mu \mathcal{H}_z^2}{2} \right) dv \end{aligned}$$

108. Boundary Conditions for Electric and Magnetic Fields. In order that any electromagnetic field distribution may be realized physically, it must satisfy Maxwell's equations. In addition it must also satisfy boundary conditions for the physical system in which the electrical effects take place.

Whenever an electromagnetic field or wave meets a boundary between two mediums of different dielectric or magnetic properties, there is a change in the fields and, in general, the wave splits into two components, one of which is reflected back into the first medium, the other is refracted into the second medium. For any two mediums, 1 and 2, the tangential components of the electric field and also of the magnetic field are equal on both sides of the dividing surface. Hence, as shown in Fig. 28, the relation exists that

$$\mathcal{E}_{1T} = \mathcal{E}_{2T} \quad \text{and} \quad \mathcal{H}_{1T} = \mathcal{H}_{2T}$$

The relationship between the normal components may be determined from Gauss's law. If the surface charge density of the surface separating the two mediums is σ , then the relation exists that

$$\mathcal{D}_{1n} - \mathcal{D}_{2n} = \sigma$$

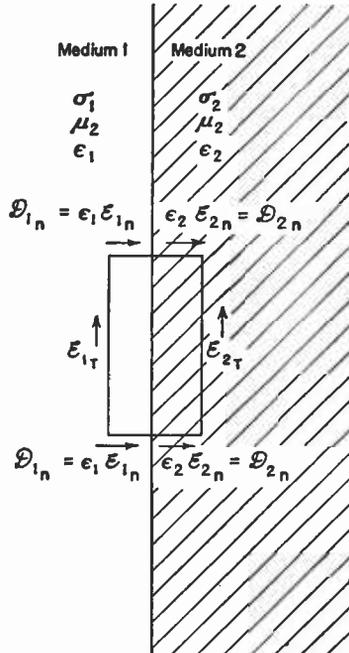


FIG. 28. Boundary conditions for electric and magnetic fields.

If, as is often the case, the surface charge density is zero, then the above relation simplifies to

$$\mathcal{D}_{1n} = \mathcal{D}_{2n}$$

From similar reasoning it can be shown that for the magnetic field, for which no free magnetic poles exist, the required relation is

$$\mathcal{B}_{1n} = \mathcal{B}_{2n}$$

References

- Abraham, M., and R. Becker: "The Classical Theory of Electricity and Magnetism," Blackie, 1932.
 Atwood, S. S.: "Electric and Magnetic Fields," Wiley, 1932.
 Bateman, H.: "Mathematical Analysis of Electrical and Optical Wave Motion on the Basis of Maxwell's Equations," Cambridge, 1915.
 Briggs, H. F.: "The Electromagnetic Field," Clarendon Press, 1934.
 Duff, A. W., and S. J. Plimpton: Elements of Electromagnetic Theory," Blakiston, 1940.
 Hund, A.: "Phenomena in High-frequency Systems," McGraw-Hill, 1936.
 Jeans, J. H.: "Mathematical Theory of Electricity and Magnetism," Cambridge, 1943.
 King, R.: "Electromagnetic Engineering," Vol. I, McGraw-Hill, 1945.
 Mason, M., and W. Weaver: "The Electromagnetic Field," University of Chicago Press, 1929.
 Maxwell, J. C.: "Treatise on Electricity and Magnetism," Oxford, 1904.
 Moullin, E. B.: "Principles of Electromagnetism," Oxford, 1932.
 Pierce, G. W.: "Electric Oscillations and Electric Waves," McGraw-Hill, 1920.
 Ramo, S., and J. R. Whinnery: "Fields and Waves in Modern Radio," Wiley, 1944.
 Schelkunoff, S. A.: "Electromagnetic Waves," Van Nostrand, 1943.
 Skilling, H. H.: "Fundamentals of Electric Waves," Wiley, 1942.
 Slater, J. C.: "Microwave Transmission," McGraw-Hill, 1942.
 Stratton, J. A.: "Electromagnetic Theory," McGraw-Hill, 1941.
 White, F.: "Electromagnetic Waves," London, 1934.

UNITS AND DIMENSIONS

109. Nature of Units. Quantities of many kinds (mechanical, thermal, chemical, physiological, etc.) are encountered in scientific and engineering work. The magnitudes of different quantities are recorded or compared with one another by selecting some magnitude as unity and expressing all others of the same quantity as a ratio to the unit quantity or magnitude. Hence, any quantity is composed of two parts: (1) a numeric which expresses the relative magnitude of the quantity in terms of the established unit magnitude, and (2) a statement designating the kind of measurement that the quantity represents. For example, the velocity of 3 m per sec contains these two parts: the numeric 3 indicates that the quantity in question has a magnitude 3 times that of unit magnitude in the system of measurement under discussion, and the kind of quantity is indicated by the measure of meters per second, having the dimensions of length divided by time.

Each quantity selected as a reference unit must be established by a physical standard of some kind, or, alternatively, it must be obtained or derived from combinations of other units that are selected or recognized as standard units. The units may be artificial or natural.

For engineering use, two general systems of mechanical units are in use in the United States. The English system of units, most frequently used, is based on the custom of common law and common and generally widespread use. The metric system of units, on the other hand, is legalized by an act of Congress of July 28, 1866. The metric system is extensively used in scientific work because of its simplicity and the ease with which decimally related quantities may be converted from one set of units to another.

110. English Units. Long usage and custom have lent a considerable weight of authority to the English system of units in the United States, in spite of the unnatural and unrational relationships that exist between the various quantities in this system. Because of the unnecessary complication of the oddly related units, the English system has found little support in scientific circles. In engineering work, which is more closely related to the everyday commercial transactions than is scientific work, the English system of units is encountered to such an extent that familiarity with it is requisite. The English system is based on the foot as the unit of length, the second as the unit of time, and the slug as the unit of mass.

111. Metric Units. The metric system, which finds extensive application throughout the scientific world, is based primarily upon the prototype standards of mass and length preserved at the International Bureau of Weights and Measures at Sèvres, France. Similar standards of length and mass are deposited in numerous national laboratories and serve to define the legal standards for the nation.

A great convenience of the metric system is that all units are expressed in decimal multiples or submultiples of the unit selected for the standard or reference value. The decimal multiples or submultiples are designated in Table 12. Combinations of these prefixes may be used to designate intermediate values; for example, millimicro may be used to designate 0.000,000,001.

112. Fundamental Units and Dimensions. All engineering quantities may be expressed in terms of a comparatively few quantities that are basic or fundamental. In mechanical units, the fundamental quantities are length, mass, time, and angle. In general, each of these dimensions has a different unit or reference standard in the metric system than in the English system. The name of the unit of measurement is usually sufficient to designate the system as well as the unit of measurement in question.

Electrical quantities are usually completely specified in terms of three mechanical units and one electrical quantity (although other combinations are possible) so that mechanical units provide a basis for understanding the fundamental properties of electrical quantities.

For quantities that are derived from, or are more complicated than, the elementary or fundamental quantities, the term "dimension" is used in a slightly different sense than its usual one. In the case of derived units, each new physical quantity gives rise to a new dimension or to a new combination of fundamental quantities to express the new physical quantity. For example, in terms of the fundamental quantities of mass, length, time, and angle, area may be expressed as the product of length and length, so that the dimensions of area are length squared, or, more compactly, $s = l^2$.

Table 12. Metric Prefixes

Name of prefix	Symbol	Multiplying factor
mega-	M	1,000,000
kilo-	k	1,000
hecto-	..	100
deka-	..	10
.....	..	1
deci-	..	0.1
centi-	c	0.01
milli-	m	0.001
micro-	μ	0.000,001

Table 13. Mechanical Units and Dimensions

Name	Symbol	Dimensions $m^A l^B t^C g^D$				Equivalent magnitudes		
		A	B	C	D	Mks units	Cgs units	English units
Length.....	l	0	1	0	0	1 m	100 cm	3.281 ft
Mass.....	m	1	0	0	0	1 kg	1,000 g	2.2046 lb
Time.....	t	0	0	1	0	1 sec	1 sec	1 sec
Angle.....	θ	0	0	0	1	1 radian	1 radian	1 radian
Area.....	s	0	2	0	0	1 sq m	10,000 sq cm	10.76 sq ft
Volume.....	V	0	3	0	0	1 cu m	10 ⁶ cu cm	35.31 cu ft
Velocity.....	v	0	1	-1	0	1 m/sec	100 cm/sec	3.281 ft/sec
Acceleration.....	a	0	1	-2	0	1 m/sec ²	100 cm/sec ²	3.281 ft/sec ²
Density.....	..	1	-2	0	0	1 kg/m ³	0.1 g/cm ³	0.2048 lb/ft ³
Force.....	f	1	1	-2	0	1 newton	10 ⁷ dynes	723.30 poundals
Pressure.....	p	1	-1	-2	0	1 newton/m ²	10 ³ dynes/cm ²	67.222 poundals/ft ²
Work (energy)...	w	1	2	-2	0	1 joule	10 ⁷ ergs	0.7376 ft-lb
Power.....	P	1	2	-3	0	1 watt	10 ⁷ ergs/sec	0.7376 ft-lb/sec
Momentum.....	mv	1	1	-1	0	1 kg-m/sec	10 ³ g-cm/sec	7.233 lb-ft/sec
Angular velocity..	..	0	0	-1	1	1 radian/sec	1 radian/sec	1 radian/sec
Angular acceleration.....	..	0	0	-2	1	1 radian/sec ²	1 radian/sec ²	1 radian/sec ²

Table 14. Electrical Units and Dimensions

Name	Symbol	Definition	Dimensions			
			$\frac{m^A}{A}$	$\frac{l^B}{B}$	$\frac{t^C}{C}$	$\frac{Q^D}{D}$
Charge.....	Q	Fundamental $f = \frac{Q_1 Q_2}{d^2}$	0	0	0	1
Surface charge density.....	σ	$\sigma = \lim_{\Delta s \rightarrow 0} \frac{\Delta Q}{\Delta s}$	0	-2	0	1
Volume charge density.....	ρ	$\rho = \lim_{\Delta V \rightarrow 0} \frac{\Delta Q}{\Delta V}$	0	-3	0	1
Electric field intensity.....	\mathcal{E}	$\mathcal{E} = \lim_{\Delta Q \rightarrow 0} \frac{\Delta f}{\Delta Q}$	1	1	-2	-1
Displacement.....	\mathfrak{D}	$\mathfrak{D} = \epsilon \mathcal{E}$	0	-2	0	1
Permittivity.....	ϵ	$\epsilon = \epsilon_0 \epsilon_r = \frac{\mathfrak{D}}{\mathcal{E}}$	-1	-3	2	2
Electric flux.....	ψ	$\psi = \int \int_s \mathfrak{D}_n ds$	0	0	0	1
Pole strength.....	M	$f = \frac{M_1 M_2}{\mu l^2}$	1	2	-1	-1
Magnetic field intensity.....	\mathfrak{H}	$\mathfrak{H} = \lim_{\Delta M \rightarrow 0} \frac{\Delta f}{\Delta M} = (NI \sin \theta) / r^2$	0	-1	-1	1
Magnetic flux density.....	\mathfrak{B}	$\mathfrak{B} = \mu \mathfrak{H}$	1	0	-1	1
Magnetic flux.....	ϕ	$\phi = \int \int_s \mathfrak{B}_n ds$	1	2	-1	-1
Permeability.....	μ	$\mu = \mu_0 \mu_r = \frac{\mathfrak{B}}{\mathfrak{H}}$	1	1	0	-2
Reluctivity.....	ν	$\nu = \frac{1}{\mu} = \frac{\mathfrak{H}}{\mathfrak{B}}$	-1	-1	0	2
Electric voltage (potential and induced electromotive force).....	E, e	$E = \int \mathcal{E} \cdot dl + \frac{\partial}{\partial t} \int \int_s \mathfrak{B}_n ds$	1	2	-2	-1
Capacitance.....	C	$C = \frac{Q}{E}$	-1	-2	2	2
Magnetomotive force.....	\mathfrak{F}	$\mathfrak{F} = \int \mathfrak{H} \cdot dl$	0	0	-1	1
Reluctance.....	\mathfrak{R}	$\mathfrak{R} = \frac{\mathfrak{F}}{\phi}$	-1	-2	0	2
Permeance.....	\mathfrak{P}	$\mathfrak{P} = \frac{1}{\mathfrak{R}}$	1	2	0	-2
Current.....	I, i	$I = \frac{dQ}{dt} + K' \frac{d\psi}{dt}$	0	0	-1	1
Current density.....	s	$s = \lim_{\Delta s \rightarrow 0} \frac{\Delta I}{\Delta s}$	0	-2	-1	1
Resistance.....	R, r	$R = \frac{E}{I}$	1	2	-1	-2
Resistivity.....	ρ	$\rho = \frac{E}{s}$	1	3	-1	-2
Conductance.....	G, g	$G = \frac{I}{E} = \frac{1}{R}$	-1	-2	1	2
Conductivity.....	γ	$\gamma = \frac{s}{E} = \frac{1}{\rho}$	-1	-3	1	2
Self-inductance.....	L	$L = \frac{\phi}{I}$	1	2	0	-2
Mutual inductance.....	M	$M_{AB} = \frac{\phi_A}{I_B}$	1	2	0	-2
Energy.....	W	$W = EI t$	1	2	-2	0
Power.....	P	$P = EI$	1	2	-3	0

Table 14. Electrical Units and Dimensions. (Continued)

Mks unit	Cgs electrostatic unit*	Cgs electromagnetic unit	Treated in Sec.
1 coulomb	$10^{-1}c (\cong 3 \times 10^9)$ statcoulombs	10^{-1} abcoulomb	30
$1 \frac{\text{coulomb}}{\text{m}^2}$	$10^{-6}c (\cong 3 \times 10^4) \frac{\text{statcoulombs}}{\text{cm}^2}$	$10^{-6} \frac{\text{abcoulomb}}{\text{cm}^2}$	33
$1 \frac{\text{coulomb}}{\text{m}^3}$	$10^{-7}c (\cong 3 \times 10^3) \frac{\text{statcoulombs}}{\text{cm}^3}$	$10^{-7} \frac{\text{abcoulombs}}{\text{cm}^3}$	33
$1 \frac{\text{volt}}{\text{m}}$	$10^6c^{-1} (\cong \frac{1}{3} \times 10^{-4}) \frac{\text{statvolt}}{\text{cm}}$	$10^6 \frac{\text{abvolt}}{\text{cm}}$	34
$K'' \frac{\text{coulomb}}{\text{m}^2}$	$K''10^{-6}c (\cong K''3 \times 10^4) \frac{\text{statvolts}}{\text{cm}}$	$K''10^{-6} \frac{\text{abvolts}}{\text{cm}}$	40
$1 \frac{\text{farad}}{\text{m}}$	$10^{-11}c^2 (\cong 9 \times 10^9)$	10^{-11}	40
$K'' \text{ coulomb}$	$K''10^{-1}c (\cong K''3 \times 10^9)$ statcoulombs	$K''10^{-1}$ abcoulombs	39
1	$10^6c^{-1} (\cong \frac{1}{3} \times 10^{-3})$	10^6 poles	48, 49
$K'' \frac{\text{amp-turn}}{\text{m}}$	$K''10^{-3}c (\cong K''3 \times 10^7)$ statoersted	$K''10^{-3}$ oersteds	50, 60
$1 \frac{\text{weber}}{\text{m}^2}$	$10^6c (\cong \frac{1}{3} \times 10^{-9}) \frac{\text{statweber}}{\text{cm}^2}$	10^6 gausscs	53
1 weber	$10^6c^{-1} (\cong \frac{1}{3} \times 10^{-2})$ statweber	10^6 maxwells	52
$1 \frac{\text{henry}}{\text{m}}$	$10^7c^{-2} (\cong \frac{1}{9} \times 10^{-12})$	$10^7 \frac{\text{gausses}}{\text{oersted}}$	55, 103
$1 \frac{\text{m}}{\text{henry}}$	$10^{-7}c^2 (\cong 9 \times 10^{12})$	$10^{-7} \frac{\text{oersted}}{\text{gauss}}$	55, 103
1 volt	$10^6c^{-1} (\cong \frac{1}{3} \times 10^{-9})$ statvolt	10^6 abvolts	37, 70, 71
1 farad	$10^{-9}c^2 (\cong 9 \times 10^{11})$ statfarads	10^{-9} farad	77
$K'' \text{ amp-turn}$	$K''10^{-1}c (\cong K''3 \times 10^9)$	$K''10^{-1}$ gilberts	67
$K'' \frac{\text{amp-turn}}{\text{weber}}$	$K''10^{-6}c^2 (\cong K''9 \times 10^{11})$	$K''10^{-6} \frac{\text{gilberts}}{\text{maxwell}}$	55, 103
$K' \frac{\text{weber}}{\text{amp-turn}}$	$K'10^6c^{-2} (\cong K'\frac{1}{9} \times 10^{-11})$	$K'10^6 \frac{\text{maxwells}}{\text{gilbert}}$	103
1 amp	$10^{-1}c (\cong 3 \times 10^9)$ statamp	10^{-1} abamp	57
$1 \frac{\text{amp}}{\text{m}^2}$	$10^{-6}c (\cong 3 \times 10^3) \frac{\text{statamp}}{\text{cm}^2}$	$10^{-6} \frac{\text{abamp}}{\text{cm}^2}$	58
1 ohm	$10^6c^{-2} (\cong \frac{1}{9} \times 10^{-11})$ statohm	10^6 abohms	76
1 ohm-m	$10^{11}c^{-2} (\cong \frac{1}{9} \times 10^9)$ statohm-cm	10^{11} abohms	63, 76
1 mho	$10^{-6}c^2 (\cong 9 \times 10^{11})$ statmhos	10^{-6} abmho	76
$1 \frac{\text{mho}}{\text{m}}$	$10^{-11}c^2 (\cong 9 \times 10^9) \frac{\text{statmhos}}{\text{cm}}$	$10^{-11} \frac{\text{abmho}}{\text{cm}}$	63, 76
1 henry	$10^6c^{-2} (\cong \frac{1}{9} \times 10^{-11})$ stathenry	10^6 abhenrys	78, 79
1 henry	$10^6c^{-2} (\cong \frac{1}{9} \times 10^{-11})$ stathenry	10^6 abhenrys	78, 80
1 joule	10^7 ergs	10^7 ergs	82, 83, 106
1 watt	$10^7 \frac{\text{ergs}}{\text{sec}}$	$10^7 \frac{\text{ergs}}{\text{sec}}$	81

* $c = 2.99796 \times 10^{10}$ cm/sec.

In another example, velocity, $v = l/t$, or the ratio of the length to the time, has the dimensions of lt^{-1} ; acceleration has the dimensions lt^{-2} , force of mlt^{-2} , and so on.

In Table 13 are given the more important mechanical units and dimensions in the mks, cgs, and English systems.

113. Electrical Units and Dimensions. Electrical energy and power, being primarily thought of as mechanical in their properties, can be expressed in terms of dimensions which are strictly mechanical and which are given in Table 14. Other electrical quantities cannot be expressed in terms of mechanical units alone but require at least one other electrical dimension for their specification. The electrical quantity selected for such dimensional analysis may be arbitrarily chosen. Proposals have been made to select, as the fourth unit, that of charge, current, resistance, or the permeability of free space, and perhaps other quantities as well, as recorded in the excellent treatment by Jauncey and Langsdorf.¹ For developing a philosophy relating all derived quantities in terms of fundamental concepts, the charge is the most logical fourth unit, although resistance is more suitable for laboratory purposes where working standards are to be maintained. Electrical quantities can be expressed in terms of more than one fundamental electrical concept, with a consequent reduction in the number of mechanical quantities that are then required as fundamental. One system that has been proposed is to omit mass as a fundamental quantity in building up a system of electrical dimensions in which the fundamental quantities would be resistance, current, length, and time. No matter what four quantities are selected as the basis of electrical dimensions, they may be expressed in metric (cgs or mks) or English units; they may also be rationalized or unrationalized depending upon whether the factor 4π (resulting from a fundamental property of space) is missing from expressions for force or potential or whether it is missing from expressions for flux, magnetomotive force, and the field equations.

Table 14 lists the more important electrical quantities together with their dimensions and method of expression in several systems of measurement. The mks system of units is regarded as basic since this system has been standardized or established by international agreement and has the further advantage that its units are, in most cases, the units of practical engineering use. The question of rationalization has not yet been definitely settled, but by using the factors of Table 6 (Sec. 35) the relations of Table 14 apply for either the rationalized or the unrationalized system.

114. Comments on Units and Dimensions. The system of units and dimensions encountered in electrical work is apt to be quite confusing because of the numerous systems of measurement and the various proposals that have been made from time to time. The mks system is the only one that meets the requirements of the scientific worker and the practical engineer alike, and since it has been adopted by international agreement, it appears destined to replace the older cgs systems. The mks system of units eliminates the confusion introduced by the constant of proportionality (numerically equal to the velocity of light in free space), which relates the cgs electrostatic and the cgs electromagnetic systems of units.

The question of electrical dimensions and units is an extensive one and cannot be adequately treated here.

References

- "American Standard Definitions of Electrical Terms," AIEE, 1942.
 Curtis, H. L.: "Electrical Measurements," McGraw-Hill, 1937.
 Gray, A.: "Absolute Measurements in Electricity and Magnetism," Macmillan, 1888.
 Hague, B.: "Alternating Current Bridge Methods," Pitman, 1943.
 Hund, A.: "High-frequency Measurements," McGraw-Hill, 1933.
 Jauncey, G. E. M., and A. S. Langsdorf: "MKS Units and Dimensions," Macmillan, 1940.
 Lanchester, F. W.: "Theory of Dimensions and Its Applications for Engineers," Crosby, Lockwood & Sons, London, 1936.
 Laws, F. A.: "Electrical Measurements," McGraw-Hill, 1938.
 "Radio Instruments and Measurements," *Nat. Bur. Standards Circ. 74*, 1937.
 Schulz, E. H., and L. T. Anderson: "Experiments in Electronics and Communication Engineering," Harper, 1943.
 Vigoureux, P., and C. E. Webb: "Principles of Electric and Magnetic Measurements," Blackie, Glasgow, 1936.
¹ Jauncey, G. E. M., and A. S. Langsdorf, "MKS Units and Dimensions," Macmillan, 1940.

CHAPTER 2

RESISTANCE

BY JESSE MARSTEN¹

1. General Concepts. In any electrical conductor or system in which there is a flow of current there is a certain amount of energy continually being lost or converted into forms not readily available for use. As far as is known at present this dissipation of energy may take one of two forms: there may be an evolution of heat or radiation of energy into space. Such energy dissipation is attributed to a property of electric conductors or systems termed "resistance."

When dealing with continuous currents, the resistance of a conductor or network, R , is adequately defined by *Ohm's law*,

$$E = iR \quad (1)$$

where E is the voltage drop across the conductor or network and i is the current through it. This assumes no back emf due to polarization or other causes. In this case the dissipation of energy takes place entirely in the form of heat generation, and the rate at which electrical energy is thus converted into heat is given by *Joule's law*,

$$P = i^2R \quad (2)$$

where P is the power or rate at which electrical energy is being dissipated in the form of heat, i is the continuous current in the circuit, and R the resistance of the circuit.

Ohm's law is insufficient to define resistance in a-c circuits. It is found experimentally that the rate at which heat is evolved in a circuit exceeds that which would be necessitated by the resistance of the circuit as determined by Ohm's law. This is due to the fact that the electromagnetic and electrostatic fields around the circuit vary with time and introduce effects which increase the losses in the circuit. Among these effects may be enumerated the following major ones:

1. *Eddy-current* losses in conductors and other masses of metals in and near the circuit.
2. *Hysteresis* losses in magnetic materials.
3. *Dielectric* losses in the insulating mediums.
4. *Absorption* of energy by neighboring conductors or circuits by induction.
5. *Radiation* of electromagnetic energy into space.
6. *Skin Effect*. Increase of conductor resistance due to nonuniform current density.

All these effects result in an increase in energy loss in the circuit over and above that given by Ohm's law. It, therefore, becomes necessary to introduce the concept of *a-c resistance* or *effective resistance*, which is defined by the more general joulean relationship,

$$P = i^2R \text{ effective} \quad (3)$$

where P is the power loss in the circuit due to all causes and i is the effective current in the circuit. Ohm's law for continuous currents follows directly from this more general definition.

2. Units of Resistance. The practical unit of resistance is the *ohm* and is defined by Ohm's law when the voltage and current are unity in the practical system. It has, however, been arbitrarily defined as the resistance at 0°C of a column of mercury having a uniform cross section, a height of 106.3 cm, and weighing 14.4521 g. Owing

¹ International Resistance Company.

to the increasing use of resistors having resistances of the order of millions of ohms, the *megohm* unit is also employed. The megohm is equal to 10^6 ohms.

3. Specific Resistance. It is found experimentally that the resistance of an electric conductor is directly proportional to its length and inversely proportional to its cross section:

$$R = \rho \frac{l}{A} \quad (4)$$

The proportionality factor ρ is called the *specific resistance* of the conductor and is a function of the material of the conductor and temperature.

From this definition of specific resistance it is apparent that any number of units may be derived for specific resistance, depending upon the units chosen for l and A . The unit generally employed in practical engineering is the *ohms per circular mil foot*, and is the resistance of a 1 ft length of the conductor having a section of 1 cir mil (diameter 1 mil, 0.001 in., for a circular conductor).

4. Volume Resistivity. If, in the above definition, l and A are both unity in the same system of units, then ρ is the resistance of a unit cube of the material and may be defined as the *volume resistivity* of the material. It should be noted that volume resistivity is not the resistance of any unit volume of the material but is specifically the resistance of unit volume measured across faces whose areas are each unity.

With a knowledge of the dimensions of a conductor and its specific resistance the resistance of the conductor to d.c. may be computed from Eq. (4). Consistent units must be employed. The resistance thus computed will be correct at the temperature for which the specific resistance applies. To obtain the resistance of the conductor at any other temperature a correction must be applied.

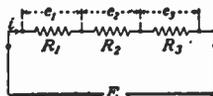


FIG. 1. Simple series circuit.

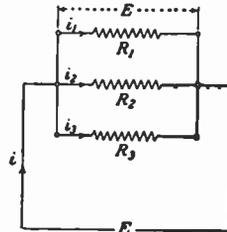


FIG. 2. Parallel circuit.

5. Temperature Coefficient. The resistance of a conductor is a function not only of the material and dimensions of the conductor but also of its temperature. Within the temperature limits generally encountered in practice, the change in resistance due to temperature variation is directly proportional to the change in temperature:

$$R_{t_2} = R_{t_1}[1 + \alpha(t_2 - t_1)] \quad (5)$$

R_{t_1} and R_{t_2} are the conductor resistances at temperatures t_1 and t_2 , respectively.

The proportionality factor α is defined as the *temperature coefficient of resistance* of the material and is the change in resistance of any material per ohm per degree rise in temperature.

All conductors do not react alike to changes in temperature. Metals, for example, have a positive temperature coefficient. Some alloys, such as manganin and constantan, have practically zero temperature coefficient and are, therefore, used primarily for resistance standards.

A knowledge of the temperature coefficient of conductor materials enables one at times to make more accurate determinations of temperature change than is possible

by thermometer measurements, especially in cases where parts to be measured are not readily accessible. Resistance determinations of the conductor are made at two temperatures and the temperature change computed from Eq. (5).

6. Properties of Materials as Conductors.

Materials	Specific resistance	Temperature coefficient
Pure metals:		
Silver.....	9.796	+0.0038
Copper.....	10.37	+0.0039
Gold.....	14.55	+0.0034
Aluminum.....	16.06	+0.00446
Tungsten.....	33.22	+0.0045
Zinc.....	35.58	+0.00347
Iron (pure).....	60.14	+0.0050
Nickel.....	60	+0.0050
Platinum.....	63.8	+0.0030
Lead.....	114.7	+0.0041
Alloys:		
"30" alloy, alloy 30 (Cu 97.75 %, Ni 2.25 %).....	30.0	+0.0013
"60" alloy, Lohm (Cu 94 %, Ni 6 %).....	60.0	+0.00071
"90" alloy, 95 alloy (Cu 89 %, Ni 11 %).....	90	+0.00049
Hytemco (Ni 72 %, Fe 28 %).....	120	+0.0045
180 Alloy, Midohm (Cu 77 %, Ni 23 %).....	180	+0.00018
Manganin (Cu 87 %, Mn 13 %).....	290	± 0.000015
Advance, alloy "45," Copel, Cupron constantan (Cu 57 %, Ni 43 %) ..	294	+0.00002
Nichrome V, alloy "A," Chromel "A," Tophet "A" (Ni 80 %, Cr 20 %).....	650	+0.00013
Nichrome, alloy "C," Chromel "C," Tophet "C" (Ni 61 %, Cr 15 %, Fe 24 %).....	675	+0.00017
Radiohm (Fe 78.5 %, Cr 16.5 %, Al 5 %).....	800	+0.0007
"331" alloy, Karma Evanohm (Ni, Cr, Al, and small trace of Fe).....	800	± 0.00002

7. Resistors in Series and Parallel. Simple and complex networks of resistors may be represented by an equivalent resistor expressed in terms of the individual resistances making up the network. The equivalent resistance of a number of resistors connected in series is equal to the sum of the individual resistances. Referring to Fig. 1,

$$E = iR_{\text{equiv}} = e_1 + e_2 + \dots + e_n = R_1i + R_2i + \dots + R_ni = i(R_1 + R_2 + \dots + R_n)$$

$$\frac{E}{i} = R_{\text{equiv}} = (R_1 + R_2 + \dots + R_n)$$

$$R_{\text{equiv}} = \sum_1^n R$$

The reciprocal of the equivalent resistance of a number of resistors connected in parallel is equal to the sum of the reciprocals of the individual resistances. Referring to Fig. 2,

$$i = i_1 + i_2 + \dots + i_n = \frac{E}{R_1} + \frac{E}{R_2} + \dots + \frac{E}{R_n}$$

$$\frac{i}{E} = \frac{1}{R_{\text{equiv}}} = \frac{1}{R_1} + \frac{1}{R_2} + \dots + \frac{1}{R_n}$$

$$\frac{1}{R_{\text{equiv}}} = \sum_1^n \frac{1}{R}$$

RESISTANCE AS FUNCTION OF FREQUENCY

8. Ratio of H-f Resistance to the D-c Resistance for Different Values of

$$x = \pi d \sqrt{2\mu f / \rho} \times \sqrt{1/1,000}$$

x	R_f/R_0	x	R_f/R_0	x	R_f/R_0
0	1.0000	5.2	2.114	14.0	5.209
0.5	1.0003	5.4	2.184	14.5	5.388
0.6	1.0007	5.6	2.254	15.0	5.562
0.7	1.0012	5.8	2.324		
0.8	1.0021	6.0	2.394	16.0	5.915
0.9	1.0034	6.2	2.463	17.0	6.268
				18.0	6.621
1.0	1.005	6.4	2.533	19.0	6.974
1.1	1.008	6.6	2.603	20.0	7.328
1.2	1.011	6.8	2.673		
1.3	1.015	7.0	2.743	21.0	7.681
1.4	1.020	7.2	2.813	22.0	8.034
1.5	1.026	7.4	2.884	23.0	8.387
				24.0	8.741
1.6	1.033	7.6	2.954	25.0	9.094
1.7	1.042	7.8	3.024		
1.8	1.052	8.0	3.094	26.0	9.447
1.9	1.064	8.2	3.165	28.0	10.15
2.0	1.078	8.4	3.235	30.0	10.86
				32.0	11.57
2.2	1.111	8.6	3.306	34.0	12.27
2.4	1.152	8.8	3.376		
2.6	1.201	9.0	3.446	36.0	12.98
2.8	1.256	9.2	3.517	38.0	13.69
3.0	1.318	9.4	3.587	40.0	14.40
				42.0	15.10
3.2	1.385	9.6	3.658	44.0	15.81
3.4	1.456	9.8	3.728		
3.6	1.529	10.0	3.799	46.0	16.52
3.8	1.603	10.5	3.975	48.0	17.22
4.0	1.678	11.0	4.151	50.0	17.93
				60.0	21.47
4.2	1.752	11.5	4.327	70.0	25.00
4.4	1.826	12.0	4.504		
4.6	1.899	12.5	4.680	80.0	28.54
4.8	1.971	13.0	4.856	90.0	32.07
5.0	2.043	13.5	5.033	100.0	35.61

It is frequently useful to know the largest diameter of wire of different materials that will give a ratio of R_f/R_0 of 1.01 for different frequencies. For a ratio of R_f/R_0 equal to 1.001, the diameters given in Sec. 10 should be multiplied by 0.55; and for R_f/R_0 equal to 1.1, the diameters should be multiplied by 1.78.

9. Skin Effect. The resistance of a conductor is a minimum when the current density is uniformly distributed over the cross section of the conductor. This condition obtains for d.c. The resistance increases for nonuniform distribution of current density over the cross section of the conductor. This latter condition obtains in conductors carrying a.c. This is a result of the distribution of magnetic flux lines, outside and inside the conductor. If the conductor is assumed to be made up of a number of conducting elements in parallel, then the interior elements, being surrounded by more flux lines than the exterior, will have greater reactance and, therefore, the current in the interior elements will be less than that in the exterior elements. As a result, the current crowds toward the surface of the conductor, giving a nonuniform current density. This imperfect penetration of current in a conductor, resulting in an increase in resistance, is termed "skin effect."

Skin effect in a conductor is a function of the following factor:

$$l \sqrt{\frac{\mu f}{\rho}} \tag{6}$$

where l = thickness of the conductor
 f = frequency of current
 μ = permeability of the conductor
 ρ = specific resistance of the conductor, mho-cm¹

It is possible to compute accurately the h-f resistance of simple round cylindrical conductors from involved functions of the above factor. To facilitate these computations, tables have been prepared from which the ratio of h-f resistance R_f to d-c resistance R_0 may be quickly determined. From this factor and the easily measured d-c resistance, the h-f resistance may be computed.

The table on page 86 gives the values of R_f/R_0 for different values of the factor

$$x = \pi d \sqrt{\frac{2\mu f}{\rho}} \sqrt{\frac{1}{1,000}}$$

where d is the diameter of the wire in centimeters, ρ is the volume resistivity in microhm-centimeters (1.724 at 10°C for copper), x may be computed for any particular case, and R_0 may be measured at d.c. or computed.

11. Reduction of Skin Effect.
 In view of the tendency of the current to crowd to the surface of the conductor at high frequencies, the remedies that have been found practical in effecting an improvement in the resistance ratio R_f/R_0 have been those in which the conductor has been designed so that it presents a thin skin to the current flow; i.e., the conductor should be a thin skin rather than a solid conductor of large cross section of which only a small part is used at high frequencies. These are

¹ Microhms per centimeter length of a conductor having a cross section of 1 sq cm.

10. Maximum Diameter of Wires for H-f Resistance Ratio of 1.01.

Frequency, kc.....	100	400	1,000	1,600	2,000	3,000	6 Mc	10 Mc	20 Mc	60 Mc	300 Mc
Wave length, m.....	3.000	750	300	187.5	150	100	50	30	15	5	1
Material	Diameter, cm										
Copper.....	0.0356	0.0177	0.0112	0.0089	0.0079	0.0065	0.00457	0.00355	0.00251	0.00145	0.00065
Silver.....	0.0345	0.0172	0.0109	0.0086	0.0077	0.0063	0.00445	0.00346	0.00267	0.00141	0.00063
Gold.....	0.0420	0.0210	0.0133	0.0105	0.0094	0.0077	0.00543	0.00422	0.00298	0.00172	0.00077
Platinum.....	0.1120	0.0560	0.0354	0.0280	0.0250	0.0205	0.01445	0.0112	0.00783	0.00456	0.00205
Mercury.....	0.264	0.132	0.0836	0.0661	0.0591	0.0483	0.03416	0.0265	0.0187	0.0108	0.00485
Manganin.....	0.1784	0.0892	0.0564	0.0446	0.0399	0.0325	0.02300	0.0179	0.0126	0.0073	0.00326
Constantan.....	0.1892	0.0946	0.0598	0.0473	0.0423	0.0345	0.02440	0.019	0.0134	0.00775	0.00346
German silver.....	0.1942	0.0970	0.0614	0.0485	0.0434	0.0354	0.02500	0.0195	0.0138	0.00784	0.00354
Graphite.....	0.765	0.383	0.242	0.191	0.171	0.140	0.0988	0.0767	0.0542	0.0312	0.0140
Carbon.....	1.60	0.801	0.506	0.400	0.358	0.292	0.2065	0.16	0.1135	0.0655	0.0292
Iron:											
$\mu = 1,000$	0.00263	0.00131	0.00083	0.00066	0.00059	0.00048	0.000339	0.000263	0.000186	0.000107	0.000048
$\mu = 500$	0.00373	0.00187	0.00118	0.00093	0.00084	0.00068	0.00049	0.000374	0.000264	0.000152	0.000068
$\mu = 100$	0.00838	0.00418	0.00264	0.00209	0.00186	0.00152	0.001075	0.000836	0.000591	0.000341	0.000153

1. *Use of Flat Copper Strip.* While skin effect is present, for the same cross-sectional area a flat strip gives a lower resistance ratio than do round conductors.
2. *Use of Tubular Conductors.* Here the external magnetic field is much greater than the internal field; therefore, all parts of the conductor are affected alike by the field, thus reducing the skin effect.
3. *Use of Litzendraht.* According to Eq. (6) the smaller the diameter of the wire, the less the skin effect. Litzendraht is a braided cable made up of a large number of fine strands of wire. When certain precautions are taken, this braid shows a very much lower resistance ratio than does a solid copper wire of equal section. These precautions are
 - a. Each strand must be thoroughly insulated from every other strand to avoid contact resistance.
 - b. Braiding must be such that each strand passes from the center to the outside of the conductor at regular intervals—a sort of transposition. This ensures that all strands are affected alike by the magnetic flux.
 - c. Each strand must be continuous.

CLASSIFICATION OF RESISTORS

Resistors generally used in radio and allied applications may be broadly classified as

1. *Fixed resistors, i.e.,* those whose ohmic value is not adjustable manually or mechanically and is intended to be substantially independent of circuit and environmental conditions, such as voltage or temperature.
2. *Variable resistors, i.e.,* those whose ohmic value is manually or mechanically adjustable, the value for any adjustment being substantially independent of circuit and environmental conditions.

NOTE: There is a class of resistors that are fixed in the sense that they are not adjustable manually and mechanically but vary with voltage or temperature by design. These are called "varistors." An example of a temperature-sensitive varistor is the so-called "thermistor" or high-negative-temperature-coefficient resistors. Examples of the voltage-sensitive varistor are the "dry-disk" rectifiers (see Sec. 39).

Each of these groups may be further classified on the basis of the conducting material used in the resistor:

1. *Wire-wound resistors,* in which metal alloy wire is the conductor.
2. *Composition resistors,* in which carbon or graphite is the principal conductor.

NOTE: There are resistors using other types of conductors; *e.g.,* resistors in which the conductor is a metallic film. The thermistor uses metallic compounds as conductor, such as nickel oxide or ferric oxide. These will be treated as special cases.

12. Fixed Wire-wound Resistors. As commonly made, wire is wound on a core of these general types: (1) textile cord or glass fiber, (2) strips of fiber or bakelite, and (3) ceramic forms. These windings are then embedded or enclosed in a coating or covering for protective purposes. The nature of the covering depends upon the core, power rating, and design of the resistor. The characteristics of the wire-wound resistor depend on the wire, core, and covering employed. Some of the performance characteristics are in large measure determined by how closely the expansion coefficients of these three elements match. In general, wire-wound resistors show a negligible or small temperature coefficient and no voltage coefficient; *i.e.,* the resistance is independent of the applied voltage.

Wire-wound resistors are used at powers ranging from less than 0.1 to 200 watts or more. To cover this wide band, different designs and structures are used, which, for convenience, give resistors that may be classified as low, medium, and high-power resistors, generally corresponding to the three core structures enumerated above. For general work these resistors are wound to accuracies of 5 and 10 per cent, although closer accuracy is possible. For precision work, a special class of resistors is generally used, which may be made to an accuracy of as low as 0.1 per cent.

13. Wire (see Sec. 6 for Constants). Two classes of alloy wire, which may be round or flat, are generally used:

1. *Nickel-copper,* in which class some of the more commonly known trade names are Advance, Cupron, Manganin, Constantan. The nickel-copper alloys are made to cover a

wide range of specific resistivities, which makes it possible to make very low resistance values without resorting to excessively large wire and resistor sizes. The temperature coefficient of these alloys, however, increases as the specific resistivity decreases. Constantan (Advance, Cupron) has a negligible temperature coefficient and is, therefore, widely used for precision resistors as well as in other types of resistors. It has a high thermal emf against copper and iron and is, therefore, widely used in thermocouples.

2. *Nickel-chromium*, in which class are the commonly known trade names of Nichrome, Tophet, and Chromel. This class of alloys has a very low temperature coefficient and the highest specific resistance of those commercially useful in resistors. As a result, it is the most generally used resistance alloy. It is made in two formulations: one iron-free, the other with a substantial percentage of iron.

Alloy wires are made in sizes as low as 0.0007 in. in diameter. However, in resistor design there is usually a limit placed on the minimum allowable size. This is usually set at 0.0013 in., the object being to reduce the possibility of failure due to corrosion and mechanical weakness. Where nickel-chromium alloy is used, it is sometimes recommended that the iron-free wire be used for wire sizes smaller than 0.002 in.

14. **Ceramic Forms.** The winding forms generally used for power wire-wound resistors are ceramic. They are also used as the insulating base for h-f composition resistors (Sec. 35) and deposited-carbon resistors (Sec. 37). Different types of ceramics are used depending upon the application, typical ones being porcelain and steatite. Some of the more important characteristics of a good ceramic for resistor use are as follows:

1. It should have high insulation resistance at room and elevated temperatures.
2. It should not have much free alkali content, otherwise ionic conduction will result with consequent electrolysis and corrosive effect.
3. It should have high mechanical strength.
4. It should have low moisture absorption, otherwise it should be protected against moisture absorption.
5. It should be capable of withstanding high thermal shock.
6. Its coefficient of thermal expansion should approximate as closely as possible that of the resistance wire and protective coating of the resistor.
7. It should have low h-f losses when the resistor is intended for high frequencies.

15. **Protective Coverings for Wire-wound Resistors.** Coverings on wire-wound resistors are employed to protect the windings from mechanical injury, to prevent electrolytic effects and consequent corrosion due to penetration of moisture, and to provide an insulating covering for the winding. Coverings most widely used in practice are as follows:

- A. Vitreous-enamel coatings
- B. Cement coatings employing inorganic binders
- C. Ceramic tube enclosures
- D. Glass or ceramic tube enclosures, hermetically sealed
- E. Cement coatings employing organic binders
- F. Molded phenolic enclosures

Coverings in the first four classifications are capable of withstanding temperatures in excess of 250°C without deterioration. They afford a high measure of protection against humidity. Exceptions to this statement are coverings employing sodium silicate (water-glass) binders which are highly hygroscopic and, therefore, unsuitable where resistance to humidity is an important factor.

The hermetically sealed tube enclosure affords the maximum protection against humidity. Such resistors are capable of standing severe thermal shock and numerous alternate immersions in salt water solution at 0 and 100°C, a requirement often specified by the U.S. Navy.

Coverings in the fifth classification are capable of withstanding temperatures up to about 175°C, varying with the nature of the binder. Resinous binders in general stand lower temperatures than asphaltic binders. They are, however, superior to the high-temperature coatings in their moisture-resistant properties. Silicone resin

binders have recently been developed which are capable of withstanding temperatures as high as 200°C and which afford a high measure of protection against humidity.

Coverings of the last classification are capable of withstanding temperatures from 100 to 160°C, depending upon the nature of the molding material used. The ordinary general-purpose molding materials with wood-flour base are good for the lower temperatures, whereas the asbestos and mica-filled materials are good for the higher temperatures.

16. Power Rating of Wire-wound Resistors. In view of the low-temperature coefficient of the resistance wires generally employed in wire-wound resistors, the resistance change with loads normally encountered is small. The power rating is, therefore, primarily determined by the power that the resistor can dissipate

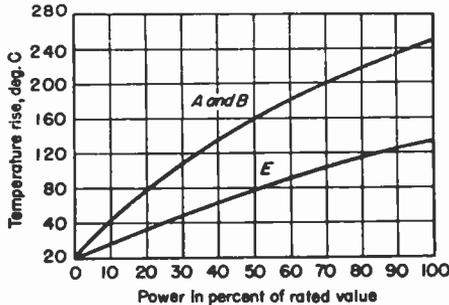


FIG. 3. Temperature rise of wire-wound resistors. *A* and *B*, vitreous enamel or inorganic cement; *E*, organic cement covering.

continuously for an unlimited time without exceeding the permissible maximum temperature of the resistor and without deterioration of the resistor. This limits the permissible temperature rise, which varies for different classes of resistors, as measured from a specified ambient temperature, usually 25 or 40°C.

with vitreous-enamel and cement coverings. The 100 per cent rating is based on manufacturers' rating of 250°C rise in open air for class A and B coatings, and 130°C rise in open air for class E coating (see Sec. 15). Temperature is measured at the center of the outer surface of the resistor, the resistor being in free air. The factors influencing this rating are

1. Heat-resistant properties of protective covering.
2. Heat-resistant properties of core.
3. Use of intermediate taps. Taps reduce effective winding space, resulting in less active cooling surface and thus reducing the nominal rating. The extent of reduction depends upon the length of the resistor, being smaller for long units than for short ones. On short units 2 in. long, the rating may be reduced by as much as 15 to 20 per cent; on units 6 in. long the reduction may be 3 to 5 per cent.
4. Ventilation. Nominal power ratings are based on specified good ventilation which is seldom encountered in practice. The actual load at which resistors may be used safely is, therefore, frequently less than the nominal rating. For example, in the case of high-power resistors nominally rated on the basis of 250°C rise, it is generally recommended that they be used in confined space at loads $\frac{1}{4}$ to $\frac{1}{2}$ nominal rating, resulting in temperature rises of 80 to 160°C. In practice, even these temperature rises may be excessive because of the proximity of the resistors to other parts that may not be subjected to elevated temperature, fire hazard, underwriters' requirements, etc. The specific application, therefore, limits the practical use of a resistor rather than any nominal rating.

18. Power Derating. In practice, the above considerations are taken into account by "derating" resistors. If a resistor is used at its nominal rating at an ambient temperature greater than its reference ambient, the maximum permissible temperature will be exceeded. To avoid this, the temperature rise must be limited by the use of power low enough to compensate the increased ambient. This is done by means of derating curves for each class of resistor showing ambient temperature vs. permissible

temperature rise, which varies for different classes of resistors, as measured from a specified ambient temperature, usually 25 or 40°C.

17. Temperature Rise of High-power Wire-wound Resistors. Although the temperature rise varies somewhat with different makes of resistors owing to differences in coating materials such as texture or color, Fig. 3 is typical and shows the temperature rise to be expected at various loadings of these resistors wound on ceramic forms,

operating load. This principle of derating applies to all classes of resistors. Every class of resistor can have such a derating curve computed from its rating and temperature rise curve. Figure 4 shows typical derating curves for fixed wire-wound resistors having class A, B, and E coatings (see Sec. 15, and Fig. 3). For derating curves applica-

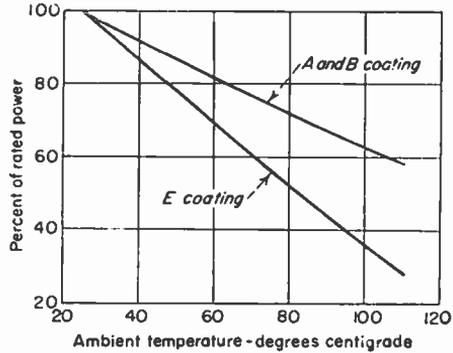


Fig. 4. Power wire-wound resistor; derating curve for high ambient temperatures.

19. Types of Fixed Wire-wound Resistors.

1. *Low-power Resistors.* These units dissipate 1/2 to 1 watt per sq in. of surface. There are two general styles: (1) flexible resistors and (2) phenolic insulated resistors. Both have a core of either cord or glass fiber.

The former has a textile or glass-fiber covering, the latter is molded in phenolic. They are equipped with wire leads making them very convenient for so-called "point-to-point" wiring in circuits, eliminating the necessity for special fittings for mounting. The flexible resistors with glass-fiber cores and coverings are capable of much higher ratings. The phenolic-molded style has become standardized in the radio industry. The table below gives the standardized data and other pertinent characteristics of this class of resistor. These units are made only in resistance values corresponding to the "preferred number" series adopted by the RMA (see Sec. 34).

Rating, watts, based on 40°C ambient	Size, in.		Min resistance, ohms	Inductance, μh	Max resistance, ohms	Inductance, μh
	Diam	Length				
1/2	3/16	5/8	0.24	0.04	820	4
1	1/4	1 1/4	0.47	0.07	5,100	52
2	5/16	1 3/4	1.0	0.1	8,200	120

2. *Medium-power Resistors (Flat Wire-wound Type).* These units dissipate between 2 and 4 watts per sq in. They consist of wire wound on strips of fiber or laminated phenolic insulation to which lug terminals are attached at appropriate points. The strip is covered with phenolic either by molding or by other means. This assembly is then tightly enclosed in a sheet-metal punching with mounting holes, or a metal mounting strip is attached in intimate engagement with one side of the resistor, enabling the other side to be mounted flat against a metal chassis.

This design has many advantages. It is easy to mount. The metal enclosure, or mounting strip, and the chassis act as heat distributors and prevent excessive differences in temperature along the length of the unit. Use is made of the metal chassis and metal mounting to conduct heat away from the resistor. This enables higher power ratings for a given temperature rise. This type of mounting in effect approximately doubles the free air rating of the resistor. For example, a resistor of this type, having a rating of 3.9 watts when suspended in free air, will have a rating of 7.4 watts when mounted on a steel plate 1 ft by 1 ft by 1/16 in.

They are used in the power range from 2 to 20 watts. Lengths vary from 2 to 6 in. Widths vary from 1/2 to 3/4 in. Terminals employed are generally soldering lugs.

3. *High-power Resistors.* These are generally single-layer inductively wound on various shapes of ceramic cores, have cement or vitreous enamel coatings, or are

hermetically sealed in glass or ceramic tubes. When inorganic cement or vitreous coatings are used, they are made to handle powers from 5 to 225 watts, depending upon the size of the unit, at dissipations of 5 to 10 watts per sq. in. of surface. These ratings correspond to a maximum operating temperature of 275°C and are based on a 250°C temperature rise above a 25°C ambient. Terminals used on these resistors are wire leads, soldering lugs, or ferrules for mounting in fuse clips.

When organic coatings are used, they are made to handle powers from 2 to 90 watts, depending upon size of unit, at dissipations of 2 to 4 watts per sq in. of surface. These ratings correspond to a maximum operating temperature of 160°C and are based on a 135°C temperature rise above a 25°C ambient.

When used at higher ambients than 25°C, they must be derated in accordance with the derating curve of the appropriate RMA or government specification (see Sec. 18).

20. Special High-power Wire-wound Resistors. 1. *Flat Type.* This is a wire-wound resistor using a rather flat elliptical ceramic with a hole of rectangular cross section through which the mounting bracket, made of a good heat-conducting metal, such as copper or aluminum, is passed. The advantage of this type is that the heat is conducted from the unit, by means of the metal mounting bracket, to the chassis on which it is mounted. This results in a more uniform temperature distribution along the length of the resistor. The mounting hardware is such that these units may be stacked one on top of another and so provide some space saving. However, the power rating of each resistor is reduced when used in this way because of the heat radiation from one unit of the stack to the other. The reduction in rating to be expected is as shown in the accompanying table.

Reduction of Rating Caused by Stack Mounting

Resistors stack-mounted on	Per cent reduction of rating when number of units stacked is		
	2	3	4
Horizontal panel.....	25	45	60
Vertical panel.....	18	30	35

2. Noninductive Type. Power wire-wound resistors of all styles can be wound noninductively by using the Ayrton-Perry winding, consisting of two windings in opposite directions, with both windings connected in parallel. This type of winding has not only low inherent inductance but low capacitance as well if wound with bare wire. The crossover turns reduce the distributed capacitance because the adjacent turns have very little potential difference between them. Typical inductance specifications for this type of resistance are shown below. Because of the double winding it is

Resistance Value, Ohms	Maximum Residual Series Inductance, Microhenrys, Measured at 1 to 3 Mc
1-99	0.5
100-999	1.0
1,000, or higher	2.0

necessary to wind the noninductive resistor with very large pitch. This limits materially the maximum value of resistance obtainable on a given form, as shown by the following examples.

Size of winding form		Max resistance value, ohms	
Diam, in.	Length, in.	Noninductively wound	Inductively wound
$\frac{5}{16}$	$1\frac{3}{4}$	12,000	2,000
$\frac{9}{16}$	3	37,000	2,000
$\frac{3}{4}$	$4\frac{1}{2}$	90,000	5,000
$1\frac{1}{8}$	$6\frac{1}{4}$	200,000	5,000

3. Adjustable Resistors. In many applications a fixed resistance value is required, but the exact value is not known. The adjustable resistor fills this need. This is a standard power resistor as described (Sec. 19) with a cement or vitreous enamel coating. A portion of the form along its entire length, usually $\frac{1}{4}$ in. wide, is left uncovered, leaving the wire exposed. This forms a track on which an adjustable terminal, which makes pressure contact on the wire, may be moved until the proper value is obtained, after which the terminal is locked in position. Such resistors are made to tolerances of ± 10 per cent only.

21. Precision Resistors. Any of the types of wire resistors previously described can be made to a close accuracy. However, they were not designed specifically for precision purposes and so do not lend themselves to this type of application. Precision resistors, as used in instrument work, gain controls, etc., are resistors having an accuracy of ± 1 per cent or less. They are made in resistance values from a fraction of an ohm to 4 megohms or more.

Precision resistors are wound on forms, usually ceramic, having an even number of sections, the sections being separated by thin barriers of the form. The winding consists of enameled resistance wire, multiple-layer wound, adjacent sections being wound in opposite directions. The minimum wire size used is 0.0015 in. over the enamel. The winding is usually continuous, although as many as four breaks, welded or brazed, are permitted in practice. Although adjacent sections are wound in opposite directions, the resistors are not noninductive. In audio circuits they do, however, help reduce inductive pickup. The wound resistor is then protected by an insulating coating to protect it against humidity. The coating may be wax, varnish, or any other suitable material. Some resistors are hermetically sealed in a ceramic container. Terminals used are wire leads or soldering lugs.

General Data. These resistors are made in a large number of sizes ranging from $\frac{3}{16}$ by $\frac{3}{16}$ in. to $\frac{1}{8}$ by $2\frac{3}{8}$ in. They are made in resistance values from 0.1 ohm to 4.0 megohms. Standard tolerances are $\frac{1}{4}$, $\frac{1}{2}$, and 1 per cent, although they can be made to ± 0.1 per cent. There has not been sufficient standardization of ratings on this type of resistor. There are two ratings, power and voltage, the rating that is first exceeded being the applicable one. Temperature rise is used as the basis of power rating, and this varies from 20 to 40°C rise.

Power ratings vary from $\frac{1}{2}$ to $1\frac{1}{2}$ watts for the different sizes, with a maximum voltage limitation of approximately 75 to 125 volts per section, depending upon the design and manufacturer. Thus a two-section resistor would have a maximum voltage limitation of 150 to 250 volts.

The power rating is the limiting rating up to the resistance value that requires the maximum voltage rating. Higher resistance values are limited by the voltage rating.

22. Variable Wire-wound Resistors. 1. *Low Power.* These are usually of the continuously variable type made by winding resistance wire on an insulating form of bakelite, fiber, or similar material, which may be arcuate in shape or a flat strip formed into an arc after winding. The form is then placed in a protective container. A metallic sliding arm is arranged to travel over the winding, thus making contact with each turn as it is rotated. The choice of wire and size is determined by the resistance and space requirements.

In general, wire-wound continuously variable resistors are wound so that the resistance changes uniformly with the motion of the sliding contact. For certain uses, e.g., antenna-type volume controls, it is desirable that the resistance change be nonuniform. In this case, the form on which the wire is wound is sometimes tapered so that the resistance per degree of rotation is not constant. Other methods of tapering employed are winding with variable pitch, winding sections of the control with different sizes of wire, and copperplating the start and finish of the winding.

The resistors are made in sizes varying from approximately $1\frac{1}{4}$ in. in diameter and $\frac{1}{2}$ in. deep to 5 in. in diameter and $2\frac{1}{2}$ in. deep. The maximum resistance varies from 10,000 ohms for the smallest size to 500,000 ohms for the largest. The standard power ratings vary from 2 watts for the smallest size to 25 watts for the largest. Usual standard ratings are 2, 3, 7, 8, 12, and 25. The power rating is based on a temperature rise of 60°C, when

mounted on a steel plate $\frac{1}{16}$ in. thick and 4 to 6 in. square. This rating is based on a resistor wound linearly.

2. *High Power.* There are two general types of high-power variable resistors, usually made in two ways:

a. *Ceramic Type.* Resistance wire is wound on a toroidal ceramic core, and the winding coated with vitreous enamel and bonded to a ceramic base or otherwise mounted on materials having high temperature resistance. A sliding contactor, which may be metallic, or a graphite-metal brush is arranged to rotate around an open track on the resistor winding.

These resistors are made in sizes varying from $1\frac{1}{2}$ to 12 in. in diameter. They are made in power ratings from 10 to as high as 1,000 watts; the usual standard ratings being 25, 50, 75, 100, 150, 225, 300, 500, 750, and 1,000 watts. Standard tolerance is ± 10 per cent. The maximum resistance values that can be wound are very high, based on the size of these rheostats. Since these units are intended for relatively low voltage systems in which heavy currents and high power are controlled, the maximum resistance values are generally limited to 5,000 to 10,000 ohms.

Power Rating. Power rating is the power dissipated in the total resistance of the rheostat. Linear-wound rheostats of this type having ratings of 100 watts and less are rated on the following basis: When mounted in the center of a 0.063 in. thick steel panel, 12 in. square and vertically suspended in still air at an ambient of 40°C , the temperature rise, measured at the center of the open track, at rated power, shall not exceed 300°C . For rheostats above this rating, under similar conditions, the temperature rise shall not exceed 350°C . Enclosed rheostats are rated at 50 per cent of their unenclosed rating.

b. *Metal-clad Type.* Resistance wire is wound on an asbestos insulated metallic strip of high heat conductivity (aluminum or copper), which is formed into an arc or circle after winding. This is assembled, with appropriate insulation, in a metallic housing, usually die-cast aluminum. A contactor, metallic or metal-graphite, is arranged to rotate around an open track on the resistor winding.

These resistors are currently made in two sizes, approximately $1\frac{1}{2}$ and $2\frac{1}{2}$ in. in diameter, rated at 25 and 50 watts, respectively. Maximum resistance values would be 5,000 and 10,000 ohms, respectively.

Power Rating. For this type of rheostat the power rating, under the same conditions as given for the ceramic type of rheostat, is based on a temperature rise of 140°C for the 25-watt rheostat, and 170°C for the 50-watt rheostat. The reason for the lower temperature rise in this type of rheostat is that the heat generated in the winding is carried away by conduction to the metallic housing and mounting panel.

Operation at Fractional Rotation. In a rheostat the active portion may be any fraction of the total winding. The permissible power in this active portion of a linearly wound rheostat is a function of

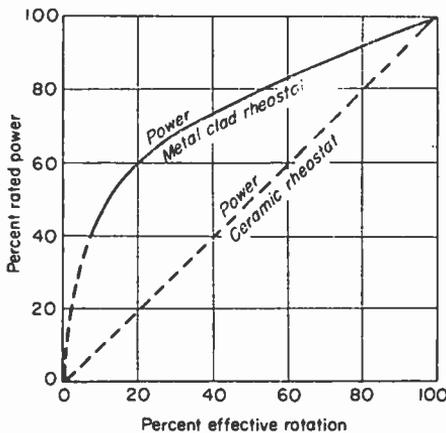


FIG. 5. Per cent rated power and current vs. per cent effective rotation. Resistor mounted on 12- by 12- by $\frac{1}{16}$ -in. steel panel.

the amount of rheostat in circuit. In the ceramic type it is practically a linear function; *i.e.*, for rated temperature rise the power varies directly with rotation. For example, at one-third rotation, one-third rated power may be applied without exceeding rated temperature rise. In the case of the metal-clad type, it is a nonlinear function. It is possible to apply very high powers at fractional rotations without exceeding permissible temperature rise. For example, at 50 per cent rotation, 80 per cent of

rated power may be applied. This is possible because of the high heat conductivity of the winding core. The heat generated in a portion of the winding core is conducted throughout the whole winding core, so that the rheostat behaves as though the entire winding were active. This is illustrated by Fig. 5.

Derating. When rheostats are operated at ambient temperatures greater than 40°C, they must be derated in accordance with specifications applicable to the particular type of unit. Figure 6 illustrates one such derating curve.

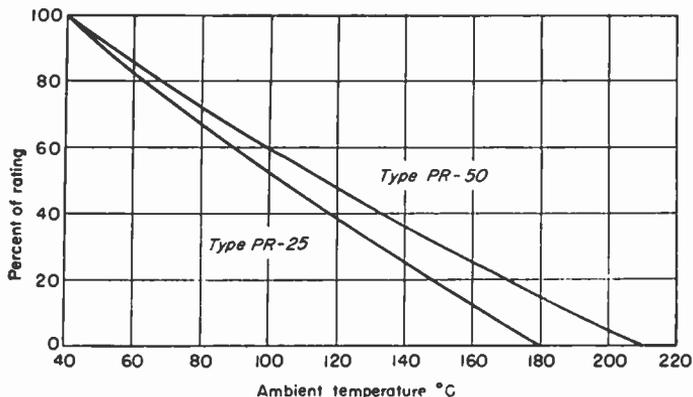


FIG. 6. Derating curves for metal-clad power rheostats, 25- and 50-watt ratings for ambient temperatures above 40°C.

Design Consideration. Some of the factors to be considered in design are as follows:

1. Contact between slider and resistor element should be positive.
2. Winding should not become loose on the form.
3. Sliding contact should not wear away resistance wire.
4. Resistance change per turn should be as small as possible.
5. Slider material should be such that it will not oxidize, wear excessively, or pit.

23. Fixed Composition-type Resistors. The term "composition-type resistor" is employed to cover that group of resistors in which a conductor is mixed with a binder in definite proportions and suitably treated to produce a resistor material. The physical form it takes may be a solid body or a film. By and large, these resistors, whether of the solid body or film type, are made in the same general way although there are wide variations in materials, formulations, details of procedure, etc. Resistors of this class are made by preparing a homogeneous mix of a conductor, generally graphite or carbonaceous, or both, with a resinous binder and sometimes filler. In the case of the solid resistor, the binder is a resinous molding powder; in the case of the film resistor, the resin is in the form of a varnish. In both cases a curing operation is essential to harden the mass or film. The mixtures have infinite or very high resistance prior to the final polymerizing operation. This operation, involving temperature or pressure, or both, shrinks and fuses the resistive composition into a tough hard body or film. The resulting structure then consists of conducting particles held together by a dielectric binder. Resistance values are controlled by modifying the ratio of dielectric to conductor. The higher this ratio, the higher the resistance value. The conducting particles make contact with each other under the pressure that exists in the structure. Most of the resistivity of the device is a summation of the contact resistances between particles, so that whatever alters the contact pressure will affect the resistivity. Many of the performance characteristics of this class of resistor are a direct result of the structure of this device.

This type of resistor is made up for different general applications as listed below.

24. Composition-type (Radio Low-power) Resistors. The composition-type resistor has attained a wide popularity in the manufacture of radio and allied products because of the following advantages: (1) flexibility in range—it may be made in any values up to several hundred megohms; (2) compactness—its physical dimensions are small for any range and they may be made in sizes as small as $\frac{3}{32}$ in. in diameter by $1\frac{3}{4}$ in. long.

Numerous styles of these resistors have been produced but they take two general forms:

1. *Solid-body Resistors.* In this style the resistor material is extruded, pressed, or molded into its final physical form, which generally is a solid rod. After this it may be subjected to some form of heat-treatment. The so-called "carbon" resistor is an example of this style. It is produced both as an uninsulated resistor (as described) and as an insulated resistor with an outside molded insulating covering. Wire leads are used as terminals.

2. *Filament-coated Resistors.* Here a conducting coat or film is baked on the surface of a continuous glass filament or other form. In the case of the glass filament this is completely enclosed in an insulating housing. The so-called "insulated-filament" resistors are examples of this style. Wire leads are used as terminals.

The insulated type of resistor, in which the resistor element is molded in a phenolic covering is used almost universally and may be considered standard in the industry.

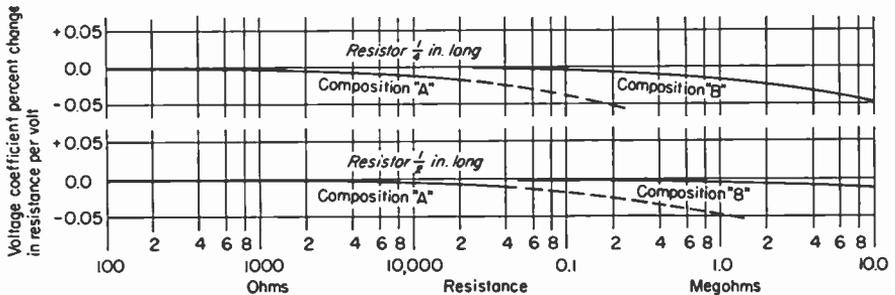


Fig. 7. Voltage characteristics of composition resistors.

Resistors of this type possess properties differing very markedly from those of metallic resistors. The most important are as follows and are possessed by all styles in varying degree.

25. Voltage Characteristics. The resistance is not independent of the applied voltage and falls with increasing voltage. This characteristic is measured by a factor called "voltage coefficient" which is defined as

$$\text{Voltage coefficient} = 100 \frac{R_1 - R_2}{R_2} \times \frac{1}{E_1 - E_2}$$

where E_1 = rated continuous working voltage

E_2 = 0.1 rated continuous working voltage

R_1 = resistance at rated continuous working voltage

R_2 = resistance at 0.1 rated continuous working voltage

The voltage coefficient is a function of

1. Composition of resistance mix.
2. Resistance value. For a given type of composition or mix and size of resistor, the coefficient decreases with decreasing resistance value.
3. Length of resistor. For a given composition and resistance value, the coefficient decreases with increasing length.

These facts are generally true of all types of composition resistors and are illustrated in Fig. 7, taken on resistors of two different lengths made with two formulas of resist-

ance material. The reason for this behavior will be clear from the following: Lower resistance values for a given composition imply less insulating material in the mix. Greater length of resistor for a given value implies lower resistance per unit length of resistor and, therefore, less insulating material. The variation required in each of these factors to produce minimum voltage coefficient is in the direction to reduce the amount of the dielectric required in the mix.

The explanation is to be found largely in the observed relationship between the contact resistance of carbon particles and the voltage across contacts. The contact resistance between two carbon particles is an inverse function of voltage. Since most of the resistance resides in the contacts, anything that reduces the voltage across the contact of conducting particles will reduce the voltage coefficient. Increasing the number of conducting particles is the equivalent of reducing the amount of dielectric in the resistor, which is the condition for minimum voltage coefficient.

26. High-frequency Characteristics—

Boella Effect. Unlike wire-wound resistors, which generally show an increase in resistance with frequency (skin effect), the composition-type resistor shows a pronounced fall-off in apparent a-c resistance from its d-c value with increasing frequency. This is particularly true of high resistance values. This effect, known as the "Boella effect" after its discoverer in Italy, was further investigated in England where G. W. O. Howe proposed a theory explaining this behavior. He assumed that the resistor simulated a transmission line of half the length of the resistor, with uniformly distributed resistance and capacitance

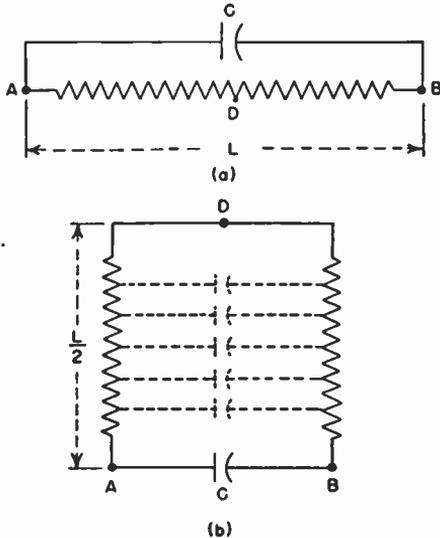


FIG. 8. (a) Composition resistor of length L . (b) Representation of (a) as transmission line having a length $L/2$ with uniformly distributed R and C .

(see Fig. 8). He derived a universal relationship between R_f/R_0 and fR_0 , in which the ratio of R_f/R_0 is constant for equal products of R_0 and f . (R_f is the apparent h-f resistance at frequency f ; R_0 equals the d-c resistance.) The fall-off in resistance value with increasing frequency is, therefore, due primarily to shunting effect of the distributed capacitance in the resistor. Many observations confirm this theory generally, although Salzburg and Miller¹ show considerable departure at frequencies above 200 Mc. More recently this subject has been studied by R. F. Field of the General Radio Company. In an unpublished paper, he introduced the additional consideration of dielectric loss in the distributed and lumped capacitances in Howe's artificial line, which would increase still further the fall-off in apparent a-c resistance. The Boella effect in composition resistors of all types is a direct result of the resistor structure. Large numbers of conducting particles interspersed with a large amount of dielectric produce capacitances and their concomitant dielectric losses, which act as shunts at high frequencies.

27. Conditions for Good H-f Characteristics. On the basis of this analysis two important conditions follow for optimum h-f characteristics:

1. *Minimum dielectric* in the resistor mass and surrounding it. The dielectric constant and loss factors should be minimum.

¹ RCA Rev., April, 1939.

2. *Geometry of Resistor.* As a direct result of this condition, the geometry of the resistor influences the h-f characteristics. Consider two resistors of equal length, one having twice the cross section of the other. For a given resistance value, the smaller unit would require resistance material with the lower specific resistivity. This means less dielectric resulting in better frequency characteristics. The condition for minimum dielectric is, therefore, minimum cross section of resistor. Similarly, if the areas of both resistors are constant, the condition for minimum dielectric is maximum length. For best h-f characteristics, the ratio of area to length of the resistor should be a minimum.

The filament-type resistor most nearly meets both requirements established for minimum Boella effect. Its cross section is extremely small since it is a film. (Film thicknesses in commercial practice vary from 0.0002 to 0.001 in.). The film volume also being small, the dielectric mass is very low.¹

Figure 9 is taken from Field.² Figure 10 represents results of an earlier set of measurements. Both sets of curves represent the average performance of a number of resistors of each type. The curves are self-explanatory and indicate the effect of geometry, insulation, and type of construction.

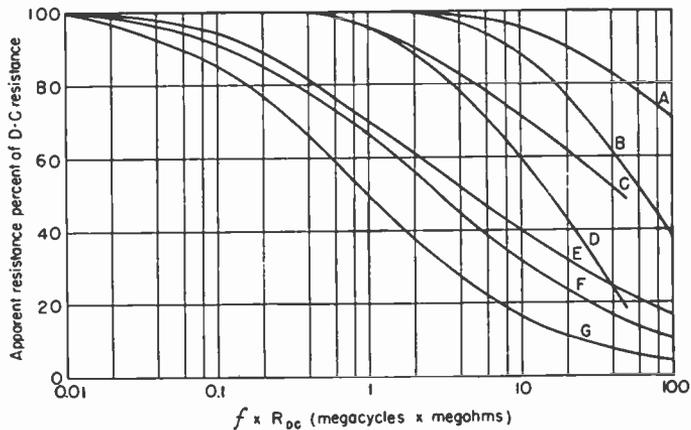


FIG. 9. Frequency characteristics of different types of composition resistors.

28. Resistor Noise Characteristics. These resistors all show, in varying degree, the presence of microphonic noise. The degree of noise is a function of the voltage, size of the resistor, resistance value, and the nature of the materials used. In general, for a given set of materials in the resistor, the noise level increases with increasing voltage, increasing resistance, and decreasing size of the resistor. Figures 11 and 12 show typical noise-level curves for two types of resistors. The points of discontinuity show where a change of mix or materials is made. The curves also show the increase in noise for a given value as the resistor size decreases. Noise measurements were made in accordance with the method described in the "Joint Army-Navy Specification on Fixed Carbon Resistors," JAN-R-11.

It is interesting to note that the pattern of factors influencing noise is similar to that influencing the voltage coefficient and h-f characteristics of composition resistors.

Christensen and Pearson³ reported that they failed to detect any noise in a solid carbon filament other than that due to thermal agitation. As a result, they concluded

¹ The most comprehensive set of h-f measurements on composition-type resistors of the solid and filament variety have been made by D. T. Drake, Massachusetts Institute of Technology, published in *Unclassified Report 520*; and by R. F. Field, previously mentioned.

² *Loc. cit.*

³ *Bell System Tech. J.*, April, 1936.

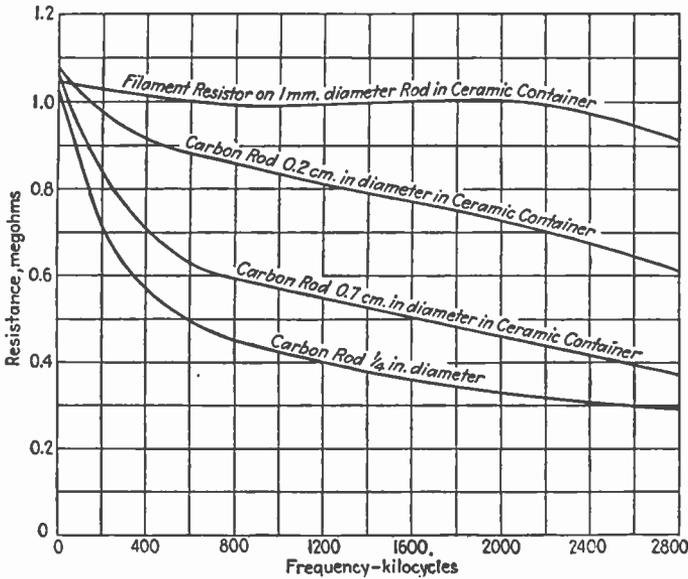


FIG. 10. Resistance-frequency characteristics of various types of 1-megohm resistors up to 3 megacycles (University of Wisconsin CWA project E-16-5).

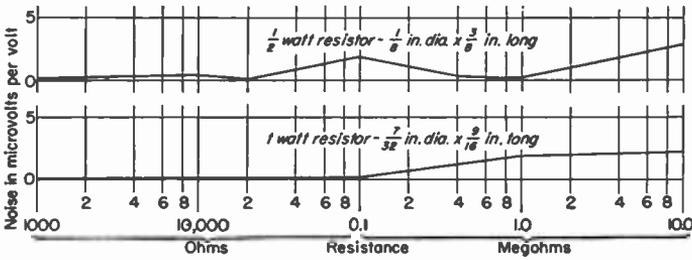


FIG. 11. Noise characteristics of composition resistors.

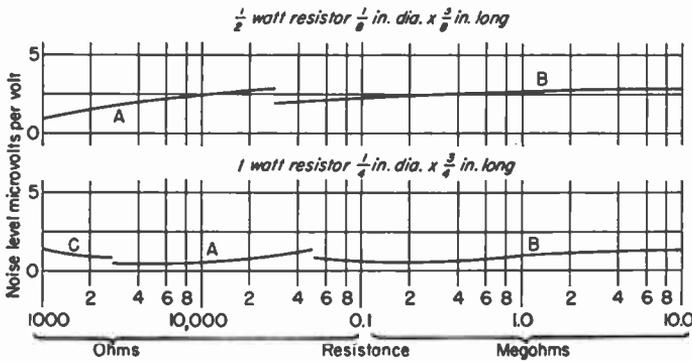


FIG. 12. Noise characteristics of composition resistors.

that most of the noise in carbon-composition resistors came from the contacts between carbon particles and that this was a function of voltage. They demonstrated that increasing the number of contacts would decrease the over-all noise. Empirical results with solid and film types of composition resistors confirm their conclusions.

In the type of resistor under discussion, resistance is increased by adding dielectric binder. This is equivalent to reducing the contact density or the number of contacts per unit volume, and increasing the resistance of the contacts. For a given current through the resistor or a given power in the resistor, this results in greater voltage drop across each contact, which produces greater noise. As in the h-f and voltage characteristic the condition for minimum noise is minimum dielectric in the resistor.

29. Humidity Characteristics. Although the general effect of humidity on resistors is to reduce resistance value because of resultant leakage, its effect on composition resistors is the opposite; resistance increases. This is because moisture causes expansion of the resistor material and separation of the conducting particles, with resulting increase in value. This effect may be reduced by proper choice of materials used and by suitable treatment of the resistor.

30. Temperature Characteristics. The temperature coefficient of this class of resistors is not a constant factor. It varies with resistance value, and for a given resistance it varies with temperature. The coefficient may even change sign at some

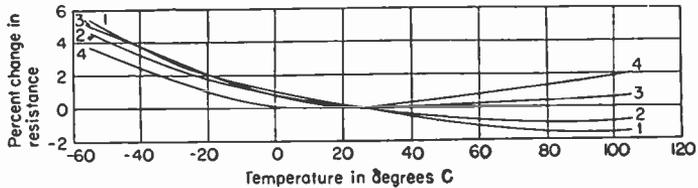


FIG. 13. Resistance-temperature characteristic, $\frac{1}{2}$ watt, $\frac{1}{8}$ by $\frac{3}{8}$ inch film composition resistor. Curve 1, 33,000 ohms; curve 2, 0.100 megohm; curve 3, 1.00 megohm; curve 4, 10.00 megohms.

point on the temperature scale. This anomalous behavior is due to the peculiar structure of the resistor. A composition resistor is one in which the conductor (graphite, carbon, or other material), which may have negative and/or positive coefficients, is mixed in varying proportions with insulation. In addition, there is an insulating supporting base for the film-type resistor. All these components react differently to temperature changes. Stresses are set up by even small differences in coefficient of expansion between these various components. Also, contraction and expansion may affect the contact pressure between conducting particles nonlinearly for the same temperature increment at different temperatures. The temperature coefficient of the conducting material itself—graphite, carbon, or other material—is not the dominant factor in establishing the temperature characteristic of the resistor. Rather it is the result of all the stresses set up by differential expansion, which affects the contact pressure between conducting particles, and this pressure determines the ultimate resistance and, therefore, temperature coefficient. As a result, it is necessary to speak of temperature characteristics over a wide range of temperatures rather than to cite a constant temperature coefficient.

Figures 13 and 14 show typical temperature characteristics of two types of composition resistors for different values and over a wide range of temperatures.

31. Rating Composition-type (Radio) Resistors. The rating of composition-type resistors is complicated by many variables. Their temperature coefficient being large, it is possible for a resistance change to become quite appreciable before a temperature limitation is exceeded. Furthermore, with the higher ranges, such as 0.25 megohm and higher, in which the power dissipation may be very low, the voltage characteristics may be a determining factor instead of the load-carrying characteristics. Also it is possible that the insulating materials used in and around the resistor may

have some influence on the rating. It is, therefore, customary to rate this type of unit on the basis of the maximum load it can carry, or the maximum voltage that can be applied to it for a long time, without exceeding prescribed resistance changes. The prescribed changes generally accepted are 10 per cent for intermittent rated-load

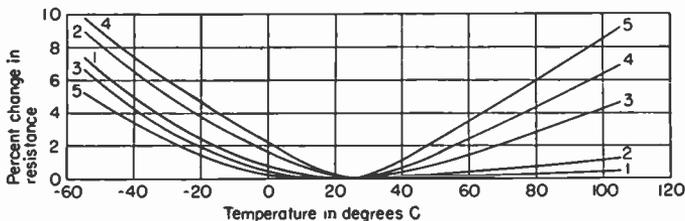


FIG. 14. Resistance-temperature characteristic, $\frac{1}{2}$ watt, $\frac{1}{8}$ by $\frac{3}{8}$, solid composition resistor. Curve 1, 1,000 ohms; curve 2, 10,000 ohms; curve 3, 0.100 megohm; curve 4, 1.00 megohm; curve 5, 10.00 megohms.

operation and 10 per cent for 50 per cent overload operation. As a result of recent developments, notably the development of insulated resistors in which the resistance element is molded in phenolic and also the development of new mixes, it has been possible to increase the rating of given sizes of resistors. This is particularly true of small resistors. This was made possible by the discovery that the temperature rise of the very short resistors is appreciably lower for a given power dissipation than would be expected from its reduced cooling surface. The reason for this is that the metal end terminals, because of shortness of the unit, cover a substantial portion of the entire resistor, are very close to the center hot section, and, therefore, cool the resistor by conducting the heat away.

Insulated Composition Resistors

Rating, watts	Max diam of resistor, in.	Max over-all length of resistor, in.
$\frac{1}{4}$	$\frac{1}{8}$	$\frac{3}{8}$
$\frac{1}{2}$	$\frac{3}{8}$ - $\frac{1}{2}$	$\frac{3}{8}$ - $\frac{1}{2}$
1	0.280	$\frac{3}{4}$
2	$\frac{1}{2}$	$1\frac{1}{4}$

The accompanying table gives the adopted standard ratings and sizes of insulated resistors, the ratings being based on an ambient temperature of 40°C.

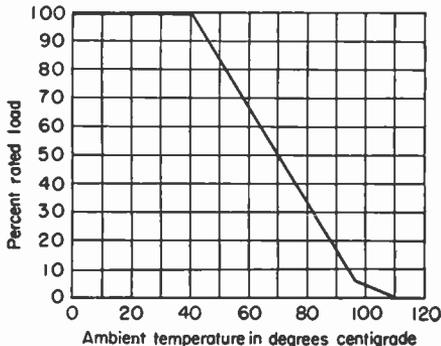


FIG. 15. Composition-resistor derating curve for high ambient temperatures.

32. Power Derating. Where composition resistors are operated at ambient temperatures in excess of 40°C, it is necessary to apply a "derating" factor to avoid exceeding permissible maximum temperatures. Figure 15 shows a typical power-derating curve generally applied to composition resistors.

33. RMA Color Code. The use of composition resistors has increased to such an extent and so many are employed in a radio set that it has become desirable to identify each resistor for range in a quick and simple manner. Such identification simplifies assembly of these units and helps in servicing. It also enables the tolerance of the resistor to be identified.

Ten colors are assigned to the figures as shown in the following table. These designations indicate the shades as shown on the Standard Color Card of America, 9th ed., 1941, issued by the Textile Color Card Association of the United States, New York.

It shall be standard in fixed-composition resistors with axial leads to indicate the nominal resistance value of the resistors by bands of color around the body of the resistor, in accordance with the following system:

Figure	Color	Color to be equivalent to cable
0	Black	
1	Brown	70118
2	Red	70179
3	Orange	70071
4	Yellow	70067
5	Green	70063
6	Blue	70209
7	Violet	70058
8	Gray	70170
9	White	

Three or more bands of color shall provide indications as follows:

Band *A* shall indicate the first significant figure of the resistance of the resistor.

Band *B* shall indicate the second significant figure.



FIG. 16. Fixed composition resistor with axial leads.

Band *C* shall indicate the decimal multiplier.

Band *D*, if any, shall indicate the tolerance limits about the nominal resistance value.

It shall be standard to indicate the significant figures of the resistance value, the decimal multiplier, and the tolerance with the modifications and extensions of the Standard RMA Color Code GEN 101 as given below:

Color	Decimal multipliers			Tolerance, per cent
	Significant figure	Power of 10	Multiplying value	
Black.....	0	10 ⁰	1	
Brown.....	1	10 ¹	10	
Red.....	2	10 ²	100	
Orange.....	3	10 ³	1,000	
Yellow.....	4	10 ⁴	10,000	
Green.....	5	10 ⁵	100,000	
Blue.....	6	10 ⁶	1,000,000	
Violet.....	7	10 ⁷	10,000,000	
Gray.....	8	10 ⁸	100,000,000	
White.....	9	10 ⁹	1,000,000,000	
Gold.....	..	10 ⁻¹	0.1	± 5
Silver.....	..	10 ⁻²	0.01	± 10
No color.....	± 20

Examples illustrating the standard are as follows:

Ohms	A	B	C
10	Brown 1	Black 0	Black, no cipher
220	Red 2	Red 2	Brown, one cipher
3,300	Orange 3	Orange 3	Red, two ciphers
4,700	Yellow 4	Violet 7	Red, two ciphers
68,000	Blue 6	Gray 8	Orange, three ciphers
150,000	Brown 1	Green 5	Yellow, four ciphers

34. Standard Resistor Values (Preferred Numbers). In the interest of standardization, the RMA has established standard resistance values in accordance with a preferred-number series. These values have been adopted universally in the industry. The series of values, appropriately rounded to two significant figures, is defined for each tolerance band as follows (RMA Standard GEN 102):

1. For tolerance of ± 5 per cent a series of values between 0.1 ohm and 100 megohms in which successive values are related by the multiplying factor, $\sqrt[24]{10}$, i.e., 1.10 approximately.
2. For tolerance of 10 per cent, a series of values between 0.1 ohm and 100 megohms in which successive values are related by the multiplying factor, $\sqrt[12]{10}$, i.e., 1.21 approximately.
3. For tolerance of 20 per cent, a series of values between 0.1 ohm and 100 megohms in which the successive values are related by the multiplying factor $\sqrt{10}$ i.e., 1.46 approximately.

The table below gives the basic decade of standard preferred resistance values for each of the three standard tolerances. These standard values apply between 0.1 ohm and 100 megohms. This standard is also applicable to fixed low-power wire-wound resistors (see Sec. 12).

Standard Preferred Resistance Values

Tolerances available, per cent					
5	10	20	5	10	20
1.0	1.0	1.0	3.3	3.3	3.3
1.1			3.6		
1.2	1.2		3.9	3.9	
1.3			4.3		
1.5	1.5	1.5	4.7	4.7	4.7
1.6			5.1		
1.8	1.8		5.6	5.6	
2.0			6.2		
2.2	2.2	2.2	6.8	6.8	6.8
2.4			7.5		
2.7	2.7		8.2	8.2	
3.0			9.1		

35. Composition-type H-f Resistors (Medium-high Power). As commonly made, these resistors consist of ceramic tubes or rods (Sec. 14) with a resistive film deposited on the outer surface of the ceramic. This film usually consists of carbon or graphite in a resinous binder that has been properly cured. Silver film terminations are applied over the ends of the ceramic, to which soldering lugs or ferrule-type terminals are applied. These resistors are made commercially in a large number of sizes. The table below gives a few typical sizes with pertinent information as supplied by the manufacturer.

Composition H-f Resistors

IRC type	Power rating, watts	Voltage rating, volts	Min available resistance, ohms	Max available resistance, megohms	Ceramic length, in.	Resistor body diam, in.	Length of resistance path, in.
MPM	$\frac{1}{4}$	200	10	1.0	$\frac{3}{16}$	0.060	$\frac{3}{8}$
MPF	2	500	20	1.5	$1\frac{1}{4}$	$\frac{3}{16}$	1
MPJ	5	1,000	35	3.0	3	$\frac{3}{16}$	2
MPP	10	1,750	25	2.5	$4\frac{1}{2}$	$\frac{3}{4}$	3.5
MPA	20	2,750	35	4.0	$6\frac{1}{2}$	$1\frac{1}{4}$	5.5
MPO	30	4,750	50	5.0	$10\frac{1}{2}$	$1\frac{1}{2}$	9.5
MPR	90	8,500	60	6.0	$18\frac{1}{2}$	2	17

Some of the characteristics are as follows:

1. The *temperature coefficient* varies with resistance value and varies from 0.005 per cent per °C for the minimum values indicated in the table to 0.07 per cent per °C for the maximum values.
2. The *voltage coefficient* also varies with resistance value and, between minimum and maximum resistance values, varies from approximately 0.002 to 0.03 per cent per volt per inch of length of resistance path. This limits the maximum voltage recommended.
3. Typical curves of the *frequency characteristics* are shown in Figs. 17, 18, and 19 for different sizes and resistance values up to a frequency of 30 Mc.
4. *Aging*. The average per cent of change per year on shelf is 3 per cent.

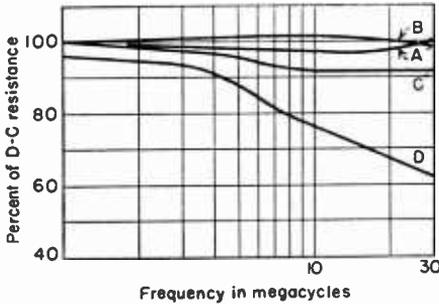


FIG. 17. Frequency characteristics of high-frequency resistors, MPM type. A, 30 to 100 ohms; B, 200 to 10,000 ohms; C, 0.01 megohm; D, 1.0 megohm.

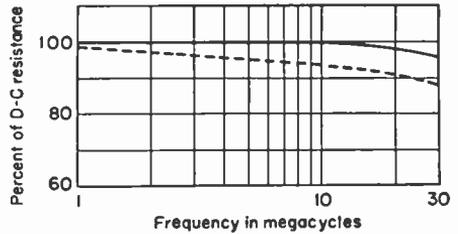


FIG. 18. Frequency characteristics of high-frequency resistors, MPF type. Full curve, 50 to 10,000 ohms; dashed curve, 0.1 megohm.

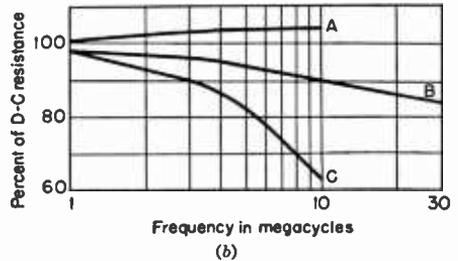
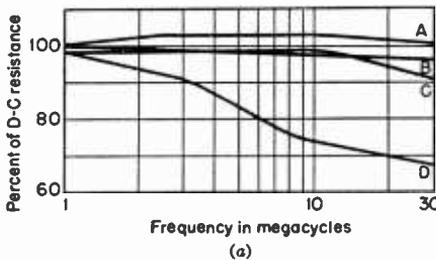


FIG. 19. Frequency characteristics of high-frequency resistors. (a) MPP resistor: A, 50 to 100 ohms; B, 1,000 ohms; C, 10,000 ohms; D, 0.1 megohm. (b) MPR resistor: A, 50 to 100 ohms; B, 1,000 ohms; C, 10,000 ohms.

Power Rating. The ratings assigned these resistors are based on operation at an ambient of 20°C and a maximum operating temperature of 90°C for continuous operation, and 110°C for short-time intermittent operation. Temperature rise curves for this type resistor are shown in Fig. 20. If operated at higher ambients, it is necessary to derate in accordance with the derating curve in Fig. 21.

Increasing Power Ratings. This type of resistor, when operated in free air, is restricted to medium power ratings because of the temperature limitations of the materials used in the resistive film. Means for increasing the power rating are

1. **Forced Air-draft Cooling.** The use of an air stream from a fan or blower can increase the rating from two to three times.
2. **Immersion in Oil.** This also has the effect of increasing the rating by a factor of 2 to 3.
3. **Water Cooling.** Extremely large increases in rating from 100 to 250 times the air rating can be obtained by appropriate methods of water cooling. This type of water-

cooled resistor is made with the resistor film on the inside of a glass tube. Water is fed through appropriate fittings tangentially into the tube at a high velocity. It flows along the resistor film in intimate thermal contact with the film and returns to its source. Resistors having a free air rating of 20 watts can be rated at 5 kw by such means. The rating depends upon the water flow. Such resistors are built so that different inlet nozzles can be installed to accommodate different water flows. The effect on rating is shown by the following table:

Effect of Water Flow on Rating of IRC Water-cooled Resistors—Type LP Having Air Rating of 20 Watts

Nozzle	Water flow, gpm	Water pressure, lb	Power rating, kw
1	4	20-25	5
2	2	15	2
3	1	15	1

36. Composition-type High-voltage Resistors. The increasing use of high-voltage sources, as in radar, television, and X-ray equipment, introduced a requirement for resistors capable of handling high voltages at moderate powers. Also these resistors must be made to extremely high values, hundreds to thousands of megohms, so that the load on the high-voltage source may be low.

High-voltage resistors are made by applying on a ceramic tube a resistance film as described in the previous section on h-f resistors, except that the resistance path is a continuous spiral line. Terminals are applied as described in Sec. 35. In this way a

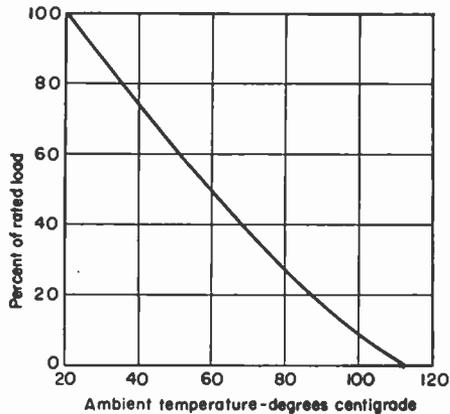
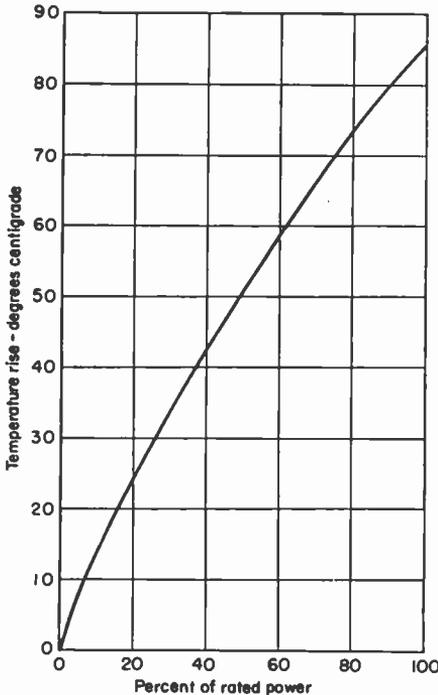


FIG. 20. Type MP resistor; load-temperature rise in free air.

FIG. 21. Type MP resistor; high-ambient-temperature derating curve.

long resistance path can be obtained on a very short tube. For example, the resistance path on a tube $4\frac{1}{2}$ in. long and $\frac{3}{4}$ in. in diameter can be made to have a linear length of 80 in. On the basis of a 250-volt-per-in. rating, such a resistor could be rated for 20,000 volts. The long resistance path also makes it possible to obtain very high resistance values more readily. For example, on the basis of 20 to 25 megohms per

in. of path, which is possible in practice, 1,500 to 2,000 megohms can be obtained on the $4\frac{1}{2}$ -in. ceramic mentioned above.

Except for the h-f characteristic, these resistors have exactly the same qualities as those of h-f resistors described above. The table below gives typical data on some of the sizes available.

High-voltage Composition Resistors

IRC type	Power rating, watts	Voltage rating, volts	Min available resistance, ohms	Max available resistance, megohms	Ceramic length, in.	Resistor body diam, in.	Length of resistance path, in.
MVF	2	5,000	2,500	250	$1\frac{3}{4}$	$\frac{3}{16}$	10
MVJ	5	10,000	20,000	1,500	3	$\frac{3}{16}$	50
MVP	10	15,000	50,000	2,000	$4\frac{1}{2}$	$\frac{3}{16}$	80
MVA	20	25,000	0.2 megohm	4,000	$6\frac{1}{2}$	$1\frac{1}{8}$	190
MVO	30	50,000	0.4 megohm	8,000	$10\frac{1}{2}$	$1\frac{1}{8}$	330
MVR	90	100,000	1.0 megohm	20,000	$18\frac{1}{2}$	2	1,100

37. Deposited-carbon Resistors. The deposited-carbon resistor is a film resistor, differing from the composition type in that the resistance film contains no dielectric material—only carbon. This film is deposited on a ceramic tube or rod (Sec. 14) by pyrolysis, *i.e.*, the cracking of a hydrocarbon gas at elevated temperatures in the proximity of the heated ceramic form. Under suitable conditions, crystalline carbon deposits on the ceramic to form an extremely hard dense film. Terminals may be wire leads, lugs, or ferrules. To obtain high resistance values, a spiral groove is cut in the carbon coating, which in effect gives a long spiral resistance line. Suitable protection is applied to the resistor in the form of a lacquer coating, phenolic tube, or hermetic sealing in glass or ceramic.

Typical Deposited-carbon Resistors Data

Manufacturer and type	Protective enclosure	Resistance range, ohms		Min tolerance per cent	Approx over-all size		Power rating, watts		Safe peak, volts
		Min	Max		Diam, in.	Length, in.	Nom	Max	
Western Electric: 140	Phenol, shell	200	10^8	± 1	$1\frac{1}{8}$ ₂	1.0	0.5	1.5	8,000
141	Phenol, shell	200	5×10^7	± 1	$1\frac{1}{8}$ ₂	$2\frac{1}{16}$	1.0	3.0	15,000
D-168753	Varnish	1	3×10^8	± 1	$1\frac{1}{16}$ ₄	$1\frac{3}{16}$	0.15	0.5	5,000
D-169025	Varnish	200	5×10^7	± 1	$\frac{3}{16}$ ₂	$2\frac{1}{16}$	1.0	3.0	15,000
D-169201	Glass	200	1.5×10^7	± 1	$1\frac{3}{8}$ ₂	$2\frac{15}{16}$	2.0	20.0	6,000
Wilkor Products: CP $\frac{1}{4}$ Watt	Varnish	20	0.5×10^8	± 1	$\frac{1}{16}$	$\frac{3}{16}$	0.25		
CP $\frac{1}{2}$ Watt	Vinylite sleeving	20	1×10^7	± 1	0.255 over sleeving	$1\frac{1}{16}$	0.5		6,000
CP 2 Watt	Vinylite sleeving	100	200×10^8	± 1	0.325 over sleeving	2	2.0	6,000
International Resistance: DCF	Varnish	200	5×10^8	± 1	$\frac{3}{16}$ ₂	$1\frac{3}{16}$	0.25	1.0	6,000
DCH	Varnish	500	2×10^7	± 1	$\frac{3}{16}$ ₂	$2\frac{1}{16}$	0.50	2.0	6,000

In effect, the physical form of this resistor is similar to the h-f resistor described in Sec. 35, and the high-voltage resistor described in Sec. 36, the important difference being that the resistance film contains no foreign dielectric material. This results in some important advantages:

1. The resistor can be operated at high temperatures, *e.g.*, at red heat, provided precautions are taken to avoid oxidation, as by hermetic sealing in vacuum. This permits high power ratings.
2. The stability of the resistor is considerably improved.
3. Aging characteristics are improved.
4. Precision is attainable because of the foregoing.

Some of the characteristics are

1. The *temperature coefficient* varies between ± 0.025 per cent per $^{\circ}\text{C}$ to ± 0.054 per cent per $^{\circ}\text{C}$.
2. The *voltage coefficient* is negligible.
3. *Aging*. The average per cent of change per year on shelf is less than 0.1 per cent.
4. Typical curve of the *frequency characteristics* is shown in Fig. 9.

The table at bottom of page 106 gives a few typical makes, and sizes and pertinent information.

38. Varistors. Varistors is the coined name applied to a class of nonohmic conductors, whose resistance, though not manually adjustable, varies over wide ranges with voltages and/or temperature. These are made of a class of materials called "semiconductors," whose resistivities at room temperature fall between those of metallic conductors and insulators. The order of magnitude of these resistivities is as follows:

Resistivity in Ohms-cm

Metals	Semiconductors	Insulators
10^{-4}	10^{-2} - 10^{-6}	10^4 - 10^{14}

These semiconductors are usually metallic oxides, or other types of metallic compounds, which may be sintered in some inert binder such as clay. They are characterized by their extreme sensitivity to voltage or temperature changes. They may be divided into three classes.

39. Low-voltage Varistors—Rectifiers. The classic examples of these are the selenium and copper-oxide rectifiers (used as power supply), and the galena or lead-sulphide and carborundum single-contact rectifiers used for detection. These are used primarily at low voltages and are characterized by their asymmetrical voltage-resistance curve; *i.e.*, the resistance is dependent on the direction of the applied voltage. They have extremely high resistance to negative voltages and extremely low resistance to positive voltages (see Fig. 22a). This property is responsible for their rectifying ability. It should be noted that this is a characteristic of single-contact varistors, *i.e.*, varistors consisting of a single crystal or disk making one contact to the other electrode.

40. High-voltage Varistors. A typical example of this class is the silicon-carbide

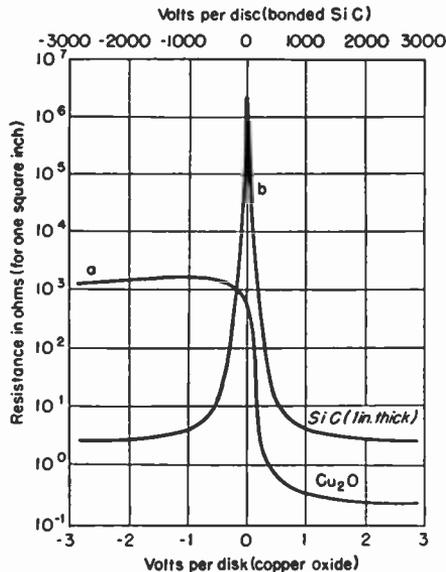


FIG. 22. Resistance-voltage characteristic of varistors: a, low-voltage copper oxide; b, high-voltage silicon-carbide varistor.

resistor (thyrite). This, in contrast to the single-contact varistor described above, consists of a large number of silicon carbide granules sintered in a vitreous ceramic. Unlike the rectifier type, the resistance voltage characteristic is symmetrical, *i.e.*, is independent of polarity (see Fig. 22*b*). This is explained partly by the fact that, although any single contact in the resistor mass may rectify, random distribution of large numbers of contacts in series and parallel result in equal numbers of contacts rectifying in opposite directions.

The characteristics of the high-voltage resistor depend upon the composition of material and heat-treatment. Changes in resistance of the order of 100,000 to 1 may be obtained by changes in the heat-treatment. Resistance changes of the order of 100,000 to 1 or more can be obtained by voltage changes of the order of 1,000 volts. Figure 23 shows a variety of characteristics possible with this type of resistor.

The temperature coefficient is negative and a function of voltage, being higher at low voltages. At low voltages a change of 45°C may change the resistance value by a factor of 2.

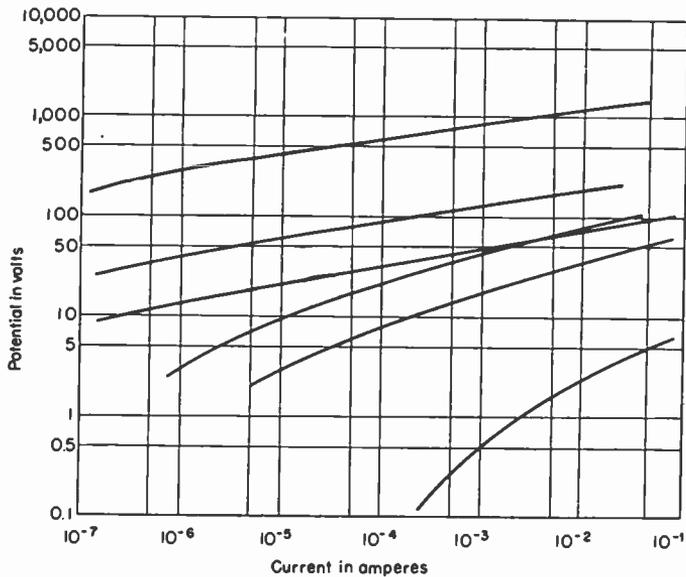


FIG. 23. Characteristics which may be obtained in silicon carbide varistors.

41. Thermistors. This class of resistors consists of mixtures of one or more metallic oxides, such as ferric oxide, Fe_2O_3 , nickel oxide, NiO , manganese dioxide, Mn_2O_2 , sintered in inert binders, such as clay or titanium dioxide. These resistors are temperature-sensitive and are characterized by an extremely high-negative-temperature coefficient. The explanation, in simplified form, is that at low temperatures, semiconductors have relatively few "free" or conduction electrons, most of them being bound to their respective atoms; hence, the resistivity is high. As the temperature is increased, however, more and more of the electrons in bound states acquire energies sufficient to carry them over into "free" energy states, where they can be accelerated by an electric field; hence, the resistance decreases, which accounts for the high negative-temperature coefficient.

There are three methods of controlling the resistance of a thermistor:

1. *External heating* or controlling ambient temperature.
2. *Direct heating*, in which the circuit current passes directly through the thermistor, thus heating the thermistor and altering its resistance.

3. *Indirect heating*, in which the thermistor is placed inside or near a heating coil that carries the control current, and the heat thus generated controls the thermistor resistance.

42. Characteristics and Uses of Thermistors.

1. *Resistance vs. Temperature.* Figure 24 shows the relationship between specific resistance and temperature for three different types of thermistor materials and for a metal,

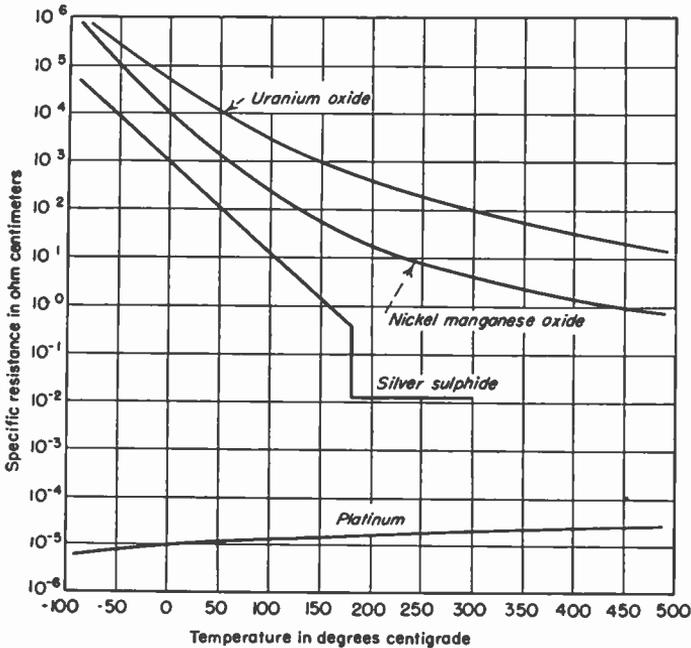


FIG. 24. Temperature characteristics of thermistor materials and a metal.

platinum. The relationship between resistance, temperature, and temperature coefficient for thermistor materials, in general, are given by

$$R = R_0 e^{\beta \left(\frac{1}{t} - \frac{1}{t_0} \right)}$$

and

$$\alpha = \frac{1}{R} \frac{dR}{dt} = - \frac{\beta}{\gamma^2}$$

where t = absolute temperature of thermistor

$R_0 = R$ at $t = t_0$

α = temperature coefficient of resistance

$e = 2.718 \dots$, the Napierian base

β = function of material and temperature

By varying the size, shape, and composition of thermistors a wide range of resistance values, and to a lesser extent temperature coefficients, can be obtained. The temperature coefficient of the resistor is also an inverse function of temperature. Figure 25 shows how the temperature coefficient of a specific type of iron-oxide thermistor varies with temperature.

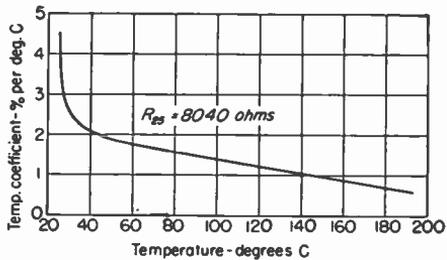


FIG. 25. Temperature coefficient vs. temperature for iron-oxide thermistor.

2. *Voltage-Current Characteristics.* Figure 26 shows the *EI* characteristic of an iron-oxide thermistor. Although the specific curve shapes vary for each type of material, they all show the same general form. An interesting feature of this curve is the falling voltage characteristic with increasing current to the right of the maximum voltage point. This shows a negative resistance characteristic, similar to that displayed by an arc, such as the Poulsen arc, indicating possible uses as an oscillator.

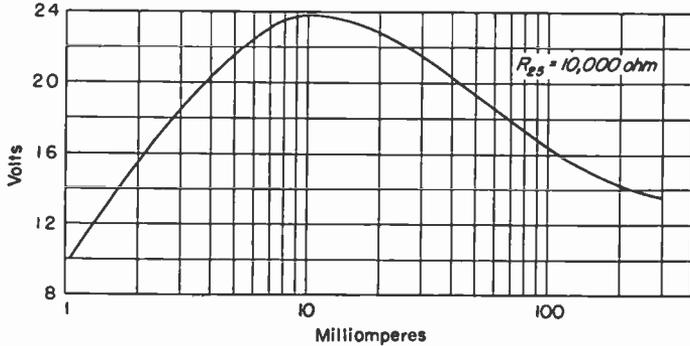


FIG. 26. Voltage-current characteristic of iron-oxide thermistor.

3. *Resistance-Power Characteristics.* Figure 27 illustrates the relationship between resistance and power in a Western Electric uranium-oxide thermistor.

4. *Current-Time Characteristics.* Owing to the thermal inertia of thermistors, the resistance change does not take place instantaneously with change in current. Figure 28 illustrates the relationship between current and time and shows the delay in current build-up.

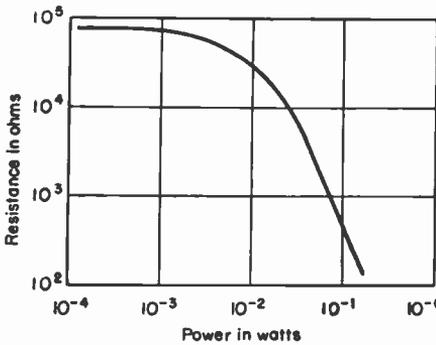


FIG. 27. Power characteristic of directly heated thermistor.

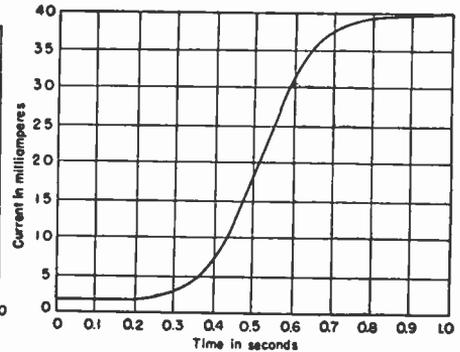


FIG. 28. Time characteristic of directly heated thermistor.

43. *Applications of Negative-temperature Coefficient Resistors.* By virtue of their high negative-temperature coefficients of resistance, their time constants, and their peculiar volt-ampere characteristics, thermistors are useful in a number of ways, a few of which are mentioned here:

1. *Resistance Thermometers.* The application here is almost obvious. A thermistor in series with an ammeter and constant-voltage source will indicate temperature when the ammeter is properly calibrated. Simple refinements of this system make it possible readily to indicate changes in temperature as small as 0.0005°C.

2. *Temperature Controllers.* Related to their use as thermometers, thermistors can be employed as temperature controllers.

3. *Time-delay Devices.* Owing to the thermal inertia of thermistors, the current through them will increase for a while after voltage is applied, as the thermistor heats up and the resistance decreases. This fact can be made use of to provide time delays ranging from a few milliseconds to several minutes.

4. *Motor Starters and Overcurrent Protection.* The high initial cold resistance of a thermistor, which decreases rapidly as it is heated by current flowing through it, is useful in motor protection or in other devices that draw high initial starting or heating-up currents.

5. *Temperature Compensation.* By proper choice and connection of thermistors the positive temperature coefficient of resistance of many passive networks can be neutralized to give a resultant network whose resistance is substantially independent of temperature over wide limits.

6. *Wattmeters.* Thermistors have been used to measure power, especially in the u-h-f band, by measuring their change in d-c resistance as they absorb all, or a given fraction, of the power to be measured.

7. *Manometers.* Since the dissipation constant of a thermistor depends on the thermal conductivity of the surrounding medium and this in turn depends, in the case of gases, on the pressure, thermistors can be readily used as sensitive-pressure gages for gases. Through a similar mechanism they can also be used as velocity meters for gases or liquids.

8. *Automatic Gain Control.* By incorporating a thermistor in the feedback circuit of a negative feedback amplifier, the amount of feedback voltage can readily be varied automatically to compensate for changes in the output signal. Many types of such AGC circuits have been devised.

9. *Amplifiers, Oscillators, and Modulators.* Since thermistors can be made to display negative resistance characteristics, they can, theoretically at least, be used as oscillators, amplifiers, and modulators. Such devices actually have been made in laboratories and have functioned well at low and audio frequencies, but there are still many difficulties that prohibit their practical use at this time.

44. *Variable Composition Resistors.* In numerous applications high-valued variable resistors are required; e.g., for controlling the sensitivity of a receiver by varying the bias on the r-f tubes a variable resistor up to 50,000 ohms maximum is commonly employed. For adjusting the audio signal level in a v-c radio sets a variable resistor up to 2.5 megohms is not uncommon. From the point of view of cost, wire-wound resistors of this magnitude are prohibitive. Furthermore, it is frequently desirable to have a nonuniform rate of change of resistance with respect to angular rotation, which is very difficult to secure with wire-wound resistors. Therefore, carbon or graphitic types of variable resistors made to meet these requirements at reasonable cost are widely used.

There are two classes of this type of resistor:

1. *Film Type.* Such resistors generally consist of a resistive solution applied to some flat form, such as paper or ceramic, and baked on. The resistor solution consists generally of carbon or graphite, or both, mixed in a resinous binder (varnish) or lacquer. This solution may be applied to the insulating base by spraying, rolling, printing, or other means of depositing a liquid. A rotating slider or some other form of contact travels over this resistive element and produces a continuous variation of resistance. Since the resistor is essentially painted on the form, its geometrical form may be varied by design. Also different concentrations of the resistor ink or paint may be employed at different positions of the resistor element. By the use of these two expedients, the resistor may be designed to give any resistance-rotation law desired.

2. *Solid Type.* In this type, the resistor element consists of carbon and graphite mixed with a resinous molding powder and molded in solid form to an insulating base. A rotating slider or other form of contact travels over this element and produces a continuous variation in resistance. By varying the ratio of conductor to binder along different positions of the resistor track, the resistor element may be made to follow different resistance-rotation laws.

In both types the resistance element is mounted on or embedded in an insulating base, and this assembly is enclosed in a metallic case. A line power switch is frequently assembled to the cover, which is operated by the shaft. Both types are made in resistance values up to 10 megohms. Standard tolerance for this type unit is 20 per cent, although they are also made to 10 per cent.

The mechanical characteristics are as follows:

1. *Size.* These controls are made in sizes varying from $\frac{3}{4}$ to $1\frac{1}{4}$ in. in diameter. Standard sizes are $1\frac{5}{16}$, $1\frac{1}{8}$, and $1\frac{1}{4}$ in.

2. *Mounting.* Single-hole mounting is standard, using a $\frac{3}{8}$ in. by 32 threads per inch as the standard mounting bushing.

3. *Rotation.* Rotation is by means of a standard $\frac{1}{4}$ in. diameter shaft, to which a knob is attached.

4. The *terminals* are solder lugs.

NOTE ON MINIATURE CONTROLS: There is a class of variable resistors, approximately $\frac{3}{4}$ in. in diameter, intended to meet space limitation requirements imposed by miniature equipment, such as hearing aids. These controls use mounting bushings or studs approximately 0.1 in. in diameter. Rotation is accomplished by a knurled cover, which encases

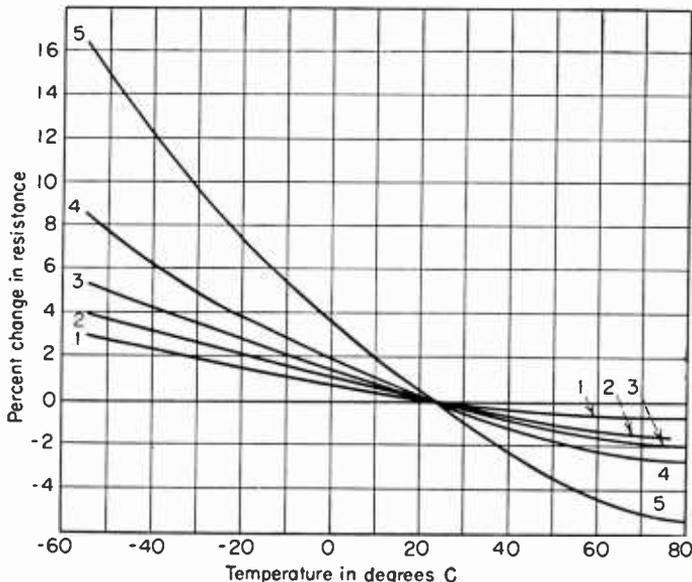


FIG. 29. Temperature-resistance characteristic; printed resistance on phenolic sheet. Curve 1, 1,000 ohms; curve 2, 10,000 ohms; curve 3, 0.100 megohm; curve 4, 1.00 megohm; curve 5, 10.0 megohms.

the control. Space occupied by the control proper is approximately $1\frac{3}{16}$ in. in diameter by $\frac{1}{4}$ in. deep.

5. *Taps.* Taps or terminals may be brought out at intermediate points along the resistance element for any purpose, *e.g.* for acoustic compensation at low levels. The standard positions for bringing such taps out are 35 and 65 per cent of effective electrical rotation.

6. *Multiple Controls.* Standard controls are made so that a number of circuits may be controlled simultaneously. Two or more single controls are mounted on one shaft and bushing assembly, so that rotation of the shaft will vary all the controls simultaneously.

7. *Concentric Dual Controls.* There are equipments, such as television receivers, in which a large number of circuits must be controlled independently. The use of individual single controls would require considerable space. So-called "concentric dual controls" are made to avoid this loss of space. These consist of a two-section control with one mounting bushing, using concentric shafts, the hollow shaft controlling the front section, the solid shaft controlling the rear section. Two circuits can be controlled independently, and the panel mounting space is that required for only one control.

The electrical characteristics are as follows:

1. *Voltage Coefficient.* Since the composition of the resistor element is the same as for fixed-composition resistors, the voltage characteristic is also the same (see Sec. 25).

2. *Temperature Characteristics.* The same considerations apply here as to fixed-composition resistors (see Sec. 30). Figure 29 shows typical characteristics for a volume-control resistance element on phenolic sheet, for different resistance values.

3. *Noise.* Standard quantitative requirements and methods of measurement have not yet been established for this characteristic. The characteristic of importance is the noise

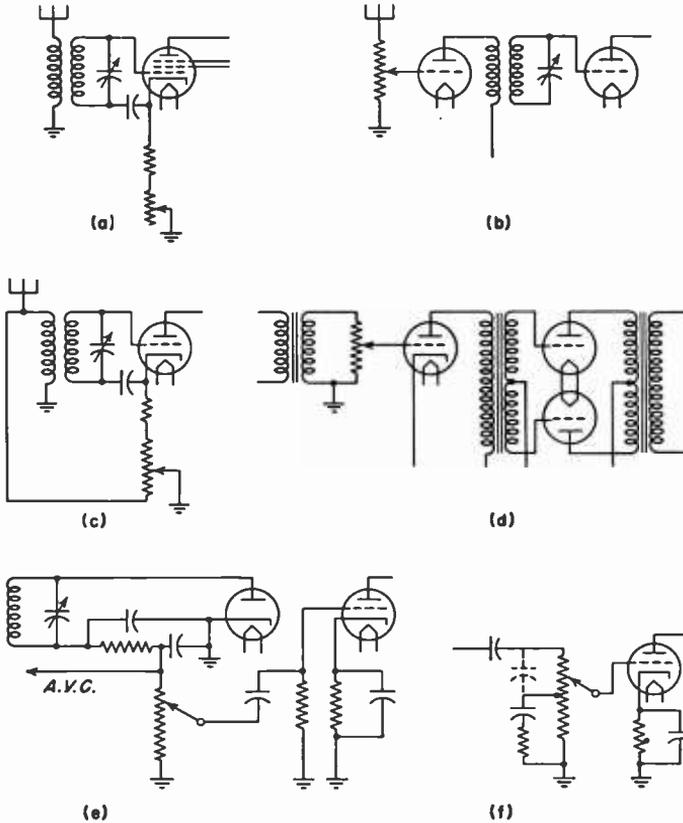


FIG. 30. Typical uses of variable resistors.

generated as a result of the movement of the contactor across the element. This is a function of a number of factors:

a. *Surface of resistor element.* A smooth flat surface produces less noise than a rough irregular one.

b. *Nature of solid contactor.* The points or area of the slider at which contact is made to the resistance element should not change or shift as the slider is rotated, if minimum noise is to be obtained. Such shifts produce irregular resistance changes and current distribution which produce noise. This imposes severe requirements on the flatness of both resistance element and contactor.

c. *Taper or resistance-rotation curve.* Smoothness and gradualness of a taper curve are important for minimum noise. Bad breaks, irregularities, and sharp changes in slope in the curve cause corresponding changes in resistance which produce noise.

d. *Presence of d.c. in the resistor circuit aggravates the production of noise.* This explains why diode load and sensitivity controls in cathode circuits used as rheostats are more noisy than audio level controls.

4. *Power Rating.* The power rating of the film-type variable resistors is $\frac{1}{4}$ watt for the $1\frac{5}{16}$ in. diameter size, $\frac{1}{3}$ watt for the $1\frac{1}{2}$ in. size, and $\frac{1}{2}$ watt for the $1\frac{1}{4}$ in. size. This

rating is increased by a factor of two to four times in some types of controls by the use of a copper heat plate between the resistance element and the mounting. The copper plate is in mechanical contact with the metal cover, bushing, and panel, so that the heat from the resistance element is carried off by the copper plate and conducted to external parts, thus reducing temperature rise.

The solid type of variable-composition resistor has a power rating of 2 watts for the $1\frac{1}{8}$ in. size, which is the only size currently made.

45. Uses for Variable Carbon Resistors.

Within their power limitation these resistors may be used wherever a continuously variable resistor is required. They may be used as either potentiometers or rheostats. They find their widest use as volume controls and tone controls in radio receivers. Some of their specific uses are here listed, and some of the basic circuits illustrating these uses are shown in Fig. 30.

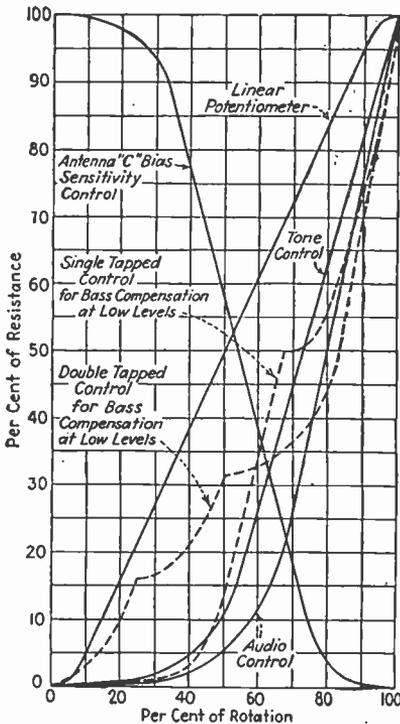


FIG. 31. Taper curves of variable resistors.

9. High-frequency variable resistor when nonreactive feature is essential, as in signal generator attenuators.

10. Television controls, such as brightness, contrast, focusing.

46. Tapers. The circuit considerations involved in these applications are discussed elsewhere in this handbook, particularly in the chapter on Receiving Systems. However, each of these applications calls for a resistance curve, or "taper" as it is termed, which is most suitable for it. This taper defines the law of resistance change vs. angular rotation of the variable arm. Some widely used curves are given in Fig. 31. A suitable specification defining the taper should include:

1. Curve showing resistance variation against active angular rotation of the contactor. Where a switch is incorporated in the variable resistor, the angle taken up for operation of the switch is considered inactive. Curve should indicate whether resistance increases with clockwise or counterclockwise rotation.

2. Resistance at extreme counterclockwise end between variable arm and left terminal; this is generally called "left terminal minimum" and is specified as "less than so many ohms."

3. Resistance at extreme clockwise end between contactor and right terminal; this is generally called "right terminal minimum" and is specified as "less than so many ohms."

4. When a tap is specified, the angular location and resistance of the tap should be given.

1. Sensitivity control for radio receivers, by varying control-grid or screen-grid potentials of r-f tubes (Fig. 30a).

2. Antenna control for varying r-f input to antenna tube (Fig. 30b).

3. Sensitivity and antenna input control. combination of Figs. 30a and 30b (Fig. 30c).

4. Audio-level control (Fig. 30d).

5. Combination load-resistor and audio-level control in diode rectifier circuit (Fig. 30e).

6. Tapped volume control for acoustic compensation at low levels. Resistance-capacitance circuits are shunted across one or more taps to produce varying degrees of a-f compensation at different levels (Fig. 30f).

7. Gain controls and faders for phonograph and a-f amplifiers.

8. Tone control in a-f amplifiers for varying a-f frequency characteristics.

The resistance between the tap terminal and the variable arm, when located at the tap, is sometimes specified.

47. Choice of Volume-control-resistance Curve. In an audio amplifier in which the maximum output is 40 db above the minimum output, the volume control should be so made that each $\frac{1}{40}$ of the rotation should correspond to an attenuation of 1 db. If the volume control has a total attenuation of 80 db, more than is necessary on this particular amplifier, each $\frac{1}{40}$ of the rotation will correspond to 2 db attenuation since only half of the total rotation can be used. In the second case the control should be more critical than in the first case.

In a radio receiver the design of the volume control differs widely depending upon whether the receiver has a v.c. or not. If not, the entire voltage gain of the receiver must be under control, perhaps 120 db. The tendency for the volume control to become noisy or to be difficult to adjust without producing violent jumps of volume change increases with the total gain that must be controlled.

The fact that a-v-c systems cannot deliver a uniform voltage to the audio detector because of the wide variations of input voltage (ranging from a microvolt to several volts) makes necessary a different shape of attenuation curve than would be used on an audio amplifier used by itself. A type of curve (Centralab Curve 6) useful in the a-v-c receiver is shown. Here, approximately uniform attenuation of 40 db is secured in 80 per cent rotation from the maximum volume. This is the range most often used. The departure from linearity in the first 15 per cent of rotation is to keep the resistance gradient within limits representing low noise.

Between 80 and 100 per cent rotation, the curve changes rapidly to provide a total attenuation of 80 db. Rapid attenuation in this region is accomplished without noise because the resistance change per decibel is small. Such a curve is much more satisfactory than a straight logarithmic line (note the 80-db curve). In addition they are simpler to build. A tapered resistance curve such that equal increments in rotation produce equal increments in attenuation (a straight line when plotted against the logarithm of the resistance) requires that a change of 300,000 ohms take place in the first 10 per cent, 120,000 ohms in the second 10 per cent, and so on till the last 10 per cent rotation produces a change of only 75 ohms. This is true of a 500,000-ohm control with a total attenuation of 80 db.

Wear Characteristics. Variable-carbon resistors necessarily have the same general electrical characteristics as fixed-carbon resistors. In addition, owing to the motion of the slider on the resistance element, there is a certain amount of wear on the resistance element. This produces a change in resistance value and noise. Factors influencing these changes are as follows:

1. Hardness of resistance element which determines ability to withstand abrasion.
2. Pressure of moving contact on resistance element.
3. Smoothness of moving contact surface.

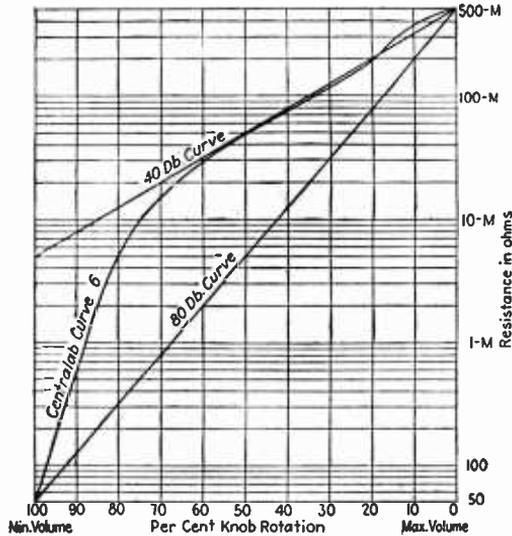


FIG. 32. Advantage of special taper for volume control.

PRINTED CIRCUITS

This name is applied to circuits where the conventional conductive wiring for interconnections is replaced by conductive metallic films.

A circuit consists of inductors (self and mutual), capacitors, and resistors interconnected by copper wiring. A printed circuit is one where any of these components is replaced by conductive or resistive films applied to an insulating base. Where this is possible, a three-dimensional device is reduced essentially to two dimensions. The art is old. There is an extensive patent literature on the application of metallic conductive films on insulation to replace conventional wiring. Inductors as well as wiring have been printed. Application of resistive films has been commercial for many years. The reduction of printing complete circuits to commercial practice is new.

48. Methods of Printing. Films, whether conductive or resistive, can be applied in several ways through stencils or shields:

1. *Spraying.* Molten metal, metal paint, resistor paints can be sprayed on the insulating base and processed.

2. *Painting.* Metallic and resistor paints can be painted or brushed on the insulation.

3. *Chemical Deposition.* Chemical solutions of metals are applied to insulation and are reduced to metal films by heating.

4. *Evaporation and Sputtering.* Metals are heated in vacuum and deposited on the insulating base by evaporation.

Printed Resistors. Printed resistors are film resistors. Methods of application and the characteristics of the film are given in Sec. 44.

RESISTOR STANDARDS AND SPECIFICATIONS

Over the past few years there has been considerable activity in the radio industry toward establishing standards and specifications for the performance of resistors and all other components. Prior to the Second World War the RMA had such a program of standards under way which would satisfy commercial requirements. The advent of the war introduced conditions of operation encountered by the military so far outside the commercial orbit that new military specifications for components were set up. For example, ambient temperatures between -50 and 100°C had to be considered. Operation at altitudes of 30,000 ft with the resultant low barometric pressure placed greater burdens on components. As a result, two sets of standards, commercial and military, for components exist. However developments in the components field are in the direction to make these two sets of standards converge.

49. Types of Tests. Performance specifications and tests have become so extensive that it is not possible to give them in detail. The following is a list of standard tests, commercial and military, to which resistors are subjected. The detailed procedures can be obtained from the specific standard specifications listed.

Standard Tests Applied to Resistors

Test	Applied to
1. Resistance measurement	All types of resistors
2. Power rating by measurement of temperature rise	All types of resistors
3. Voltage coefficient	Fixed- and variable-composition resistors
4. Temperature—resistance characteristics between -55 and 105°C	Fixed- and variable-composition resistors
5. Rapid temperature cycling between -55 and $+85^{\circ}\text{C}$	Composition resistors
6. Dielectric strength or voltage breakdown	All resistors
7. Vibration	All resistors
8. Noise †	Fixed- and variable-composition resistors
9. Life test at rated load and 40°C ambient	All resistors
10. Short-time overload	All resistors
11. Load test at high ambient temperature, 85°C	Fixed-composition resistors
12. Humidity—90 to 95 per cent relative humidity at 40°C .	All resistors
13. Salt-water cycling	Fixed-composition and wire-wound resistors
14. Thermal shock	Fixed-power wire-wound resistors
15. Wear test	All variable resistors

50. List of Specifications. The industry or commercial specifications are prepared by the engineering division of the RMA and can be obtained from the RMA Data Bureau, New York, 6. The military specifications for resistors are the Joint Army-Navy Specifications (abbreviated JAN) prepared by the Army-Navy Electronic and Electrical Standards Agency, Fort Monmouth, N.J.

Description of Standard Specifications for Resistors

	Commercial, RMA	Military, JAN
Fixed-composition resistor.....	REC-116	R-11
Fixed low-power wire-wound resistor.....	REC-117	R-184
Fixed high-power wire-wound resistor.....		R-26
Fixed precision resistor.....		R-93
Variable low-power wire-wound resistor.....		R-19
Variable high-power wire-wound resistor.....		R-22
Variable composition resistor.....	REC-121	R-94

References

Becker, Green, and Pearson: Properties and Uses of Thermistors, *Elec. Eng.*, November, 1946.
 "Deposited Carbon Resistors." Bulletin, Western Electric Company.
 Drake, D. T.: "H. F. Characteristics of Resistors," *Unclassified Report* 520, M.I.T. Radiation Laboratory.
 Driver-Harris Company: *Bull.* R-40.
 Field, R. F.: "Boella Effect," Unpublished paper.
 Howe, G. W. O.: Editorial, *Wireless Engr.*, April, 1934.
 International Resistance Company: Catalogue.
 Kennelly, Laws, and Pierce: Experimental Research in Skin Effect in Conductors. *Trans. AIEE*, 1915.
 Marsten, Jesse, and A. L. Pugh, Jr.: "Printed Resistors." Symposium Paper on Printed Circuits, National Bureau of Standards.
 Morecroft, J. H.: "Principles of Radio Communication," Wiley.
Nat. Bur. Standards Circ. 74.
 Pender, J. H.: Handbook for Electrical Engineers," Wiley.
 "Printed Circuit Techniques," *Nat. Bur. Standards Circ.* 468.
 "Resistance Handbook," Wilbur B. Driver Company.
 Resistor Standards, RMA Standards, JAN-R Standards.
 RMA Standards and Engineering Information.
 "Thermistor," Bulletin. Western Electric Company.
 "Varistors," articles reprinted from *Bell Lab. Record*.
 Wheeler, H. A.: Formulas for the Skin Effect, *Proc. IRE*, 30, 412-424, September, 1942.
 Whinnery, J. R.: Skin Effect, *Electronics*, p. 44, February, 1942.

CHAPTER 3

INDUCTANCE; MAGNETIC MATERIALS

BY F. G. WEBBER¹ AND RAYMOND L. SANFORD²

1. Definitions and Units. Inductance is that property of an electrical circuit whereby changes in the current flowing through it produce changes in the magnetic field associated with the current such that a counter emf is set up in that same circuit or in neighboring ones. If the emf is set up in the same circuit, the phenomenon is called "self-inductance"; if in a neighboring circuit, it is called "mutual inductance." Inductance is an effect exactly as inertia is an effect in a mechanical system. Inertia is that fundamental property of matter which causes a body to resist a change in its state of rest or motion. The important points are (1) that a change is necessary and (2) that the effect is to resist the change.

Inductance is measured by the emf produced in a conductor by unit rate of variation of the current. The practical unit of inductance is the *henry* (h). The henry is that inductance in which an induced emf of 1 volt is produced when the inducing current is changed at the rate of 1 amp per sec.

Since the henry is a rather large unit, especially in radio engineering work, it is commonly subdivided into two smaller units: a one-thousandth part, the millihenry (mh) = 10^{-3} henry; and a one-millionth part, the microhenry (μ h), = 10^{-6} henry.

2. The Inductor. A straight conductor carrying a current has the property of inductance and, as will later be shown, is sometimes used as an inductor. However, in other configurations this property can be enhanced to become more useful. The most common inductors are coiled conductors. Since inductance is a magnetic effect, coiling the conductors gives a more concentrated field pattern. Flux set up by one turn of the coil links mutually with other turns of the coil, thereby increasing the total inductive effect. It follows that the closer the turns, the greater is the inductance. The design of an inductor would seem to be a relatively simple matter. However, it is not possible to construct an inductor capable of pure and simple inductance. (1) There is no perfect conductor and, therefore, losses in the form of resistance are introduced. (2) Distributed capacitance exists between turns of a coil and between the conductor and surrounding objects. The actual inductor is always a combination of resistance, inductance, and capacitance. The true inductance is the absolute or theoretical inductance. The apparent inductance is the effective inductance at any frequency, *i.e.*, the inductive minus the capacitive effect.

3. Resistance Effects. The reason for the difference between the d-c and the r-f resistance of a conductor is the so-called "skin effect." Consider two small filaments of a conductor, one filament at the center and one near the surface of the conductor. All the flux produced by the current through the conductor will link the center filament, but the flux within the conductor will not link the filament near the surface. This means that the center filament will have a higher self-inductance than the surface filament. This higher self-inductance means a higher reactance path for the current in the conductor. More current, therefore, flows at the conductor surface. The redistribution of the current forces more current through less of the conductor, giving higher losses and more heating. The r-f resistance can become several times the d-c resistance. See Sec. 9, Chap. 2 for ratios of a-c to d-c resistance.

¹ F. W. Sickles Co.

² National Bureau of Standards.

Since the smaller the conductor, the less the skin effect, if the conductor is broken down into many small strands, twisted about ten turns per foot and each strand insulated from the others, the skin effect is lessened. Such wire is known as "litz" wire. Skin effect is a function of the following factors:

$$d \sqrt{\frac{\mu f}{\rho}}$$

- where d = diameter of conductor
- μ = permeability of conductor
- ρ = specific resistance of conductor
- f = frequency

For air-core inductors, skin effect is the major consideration; for iron-cored or copper-cored inductors eddy-current and hysteresis losses in the core reflect as resistances in the coil. Both these losses vary with frequency.

The eddy-current loss for any material may be expressed by

$$W_e = eVt^2f^2B_{max}^2$$

- where e = loss per unit volume
- V = volume
- t = thickness of laminations or particle size
- B_{max} = maximum flux density
- f = frequency

The hysteresis loss for iron, according to the Steinmetz empirical law, is

$$W_h = n f V B_{max}^{1.6}$$

- where n = loss per unit volume
- V = volume
- B_{max} = maximum flux density
- f = frequency

The total loss expressed in watts in an iron-cored inductor is

$$W = I^2R_{ac} + W_h + W_e$$

Another factor affecting the resistance of an inductor comes into play when that inductor "couples" with other circuit elements. If current is induced in the coupled circuit, losses occur and the resistance is increased. The general definition of effective, equivalent, or apparent resistance is such that this resistance multiplied by the square of the effective value of the current gives the true average rate at which heat energy is dissipated.

4. Distributed Capacitance Effects. The distributed capacitance can be assumed to be a capacitance shunted across the inductance. The effect is to give an apparent inductance that is less than the true inductance. The higher the distributed capacitance, the greater is the ratio of true to apparent inductance. Furthermore, because this distributed capacitance has dielectrics other than air in many cases, other losses can be introduced as dielectric losses. See the chapter on Capacitance for a further explanation of such losses.

The distributed capacitance of a multilayer coil, *e.g.*, a universal winding, is a highly complex thing. Capacitances occur between layers as well as between turns. The result is a network of capacitors in series, in parallel, and in series-parallel, which may cause the total distributed capacitance to increase or decrease as the inductance increases. This effect is shown in Table 2 (page 126), which gives the approximate distributed capacitance of choke coils. This also explains one reason why the Q of a coil may increase with inductance up to a point and then begin to decrease.

The distributed capacitance of a single-layer solenoid is a simpler function and is approximately proportional to the radius of the form and independent of the number of turns and the winding length for solenoids of the same inductance.

5. Figure of Merit or Q . To judge the quality of an inductor, which of necessity cannot be perfect, a figure of merit is used. This figure of merit, or Q , gives an idea

of the ability of an inductor to perform its primary function, that of supplying inductance at some given frequency. Q is defined as the ratio of inductive reactance to effective resistance. Therefore, the higher the true inductance, the lower the distributed capacitance, and the lower the losses from all sources, the higher the Q . Q varies with frequency since both the reactance and the losses vary with frequency.

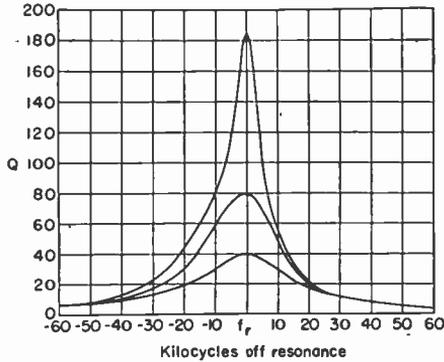


FIG. 1. Effect of Q on sharpness of resonance.

power factor will be within $\frac{1}{2}$ per cent of the reciprocal of the Q .

Q has another significance, that of showing the relative sharpness of resonance of series-resonant circuits. If the absolute value of the admittance of the circuit is plotted against frequency, a peaked resonance curve is obtained. If the resistance is increased, the admittance will be decreased at all frequencies but much more markedly near the resonant frequency. The curve will be much flatter under these conditions. The less the resistance, the steeper and sharper the curve will be, resulting in a more selective circuit. Since the inductor has the greatest part of the resistance of the circuit, the Q of the inductor is a good indicator of the expected selectivity of the circuit. High Q means good selectivity.

The curves in Fig. 1 show the sharpness of resonance for both good and poor Q windings. The curves in Fig. 2 show the Q vs. frequency characteristics for several types of windings.

MEASUREMENTS OF INDUCTANCE

The theory of the various methods of measurements to be discussed here is more thoroughly covered in the chapter on Electrical Measurements. However, the practical problems involved in measurements concerned with inductors will be considered.

6. Apparent and True Inductance. Apparent inductance may be measured by finding the capacitance necessary to resonate the inductor to a given frequency. Such measurements are frequently made since standard calibrated capacitors and frequency sources are usually handy in the laboratory. The apparent inductance is

Power factor is the ratio of impedance to resistance for any circuit. Since, for all practical inductors, the impedance and the reactance are nearly the same, the power factor is very nearly the reciprocal of Q .

$$Q = \frac{X}{R} = \tan \phi$$

$$\text{Power factor} = \frac{R}{Z} = \cos \phi$$

where ϕ is the angle between Z and R in a right-angled triangle with a hypotenuse of Z . For very large angles, the cosine is approximately equal to the reciprocal of the tangent, the cotangent. If the Q of the inductor is 10 or more, its

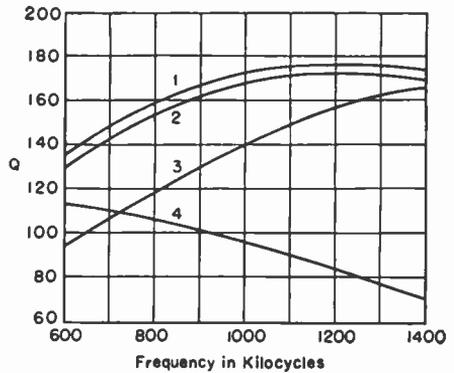


FIG. 2. Q vs. frequency for various types of windings: curve 1: two-layer bank winding; curve 2: progressive universal winding; curve 3: solenoid winding; curve 4: universal winding. All coils are 250 μ h, wound with $\frac{1}{41}$ SSE wire on $\frac{3}{8}$ -in.-OD phenolic forms.

then calculated from the formula

$$L = \frac{1}{4\pi^2 f^2 C}$$

where L is in henrys, f in cycles, C in farads, or

$$L = \frac{2.533 \times 10^4}{f^2 C}$$

where L is in microhenrys, f in megacycles, C in micromicrofarads.

Apparent inductance matching is used in the production testing of inductors when the beat-frequency or single-frequency method of testing is employed. In this case the inductor under test is compared with a standard or set-tested inductor by comparing the amount of capacitance required to tune each inductor to the same frequency.

True inductance measurements are approximated on the 1,000-cycle impedance bridge since, for most inductors used at radio frequencies, the reactance of the distributed capacitance is very high at 1,000 cycles and in parallel with the comparatively low inductive reactance it has little effect. The effective resistance is taken care of in the bridge balance.

True inductance may also be measured by finding the amounts of capacitance required to resonate the inductor to two different frequencies and then calculating the inductance by the following formula:

$$L = \frac{n^2 - 1}{4\pi^2 f_1^2 (C_2 - C_1)}$$

where f_1 and f_2 = the two frequencies

$$n = f_1/f_2$$

C_1 and C_2 = capacitances required to resonate at f_1 and f_2

The effect of distributed capacitance is eliminated as is shown by the derivation of the general true inductance formula:

$$f_1 = \frac{1}{2\pi \sqrt{L(C_1 + C_d)}} \tag{1}$$

$$f_2 = \frac{1}{2\pi \sqrt{L(C_2 + C_d)}} \tag{2}$$

$$n = \frac{f_1}{f_2} \quad \text{or} \quad f_1 = n f_2 \tag{3}$$

Substituting Eq. (3) in Eq. (1),

$$n^2 f_2^2 4\pi^2 = \frac{1}{L(C_1 + C_d)} \tag{4}$$

Substituting Eq. (2) in Eq. (4),

$$\frac{n^2}{L(C_2 + C_d)} = \frac{1}{L(C_1 + C_d)}$$

whence,

$$C_d = \frac{C_2 - n^2 C_1}{n^2 - 1} \tag{5}$$

Substituting Eq. (5) in Eq. (1),

$$f_1^2 4\pi^2 = \frac{1}{L \left(C_1 + \frac{C_2 - n^2 C_1}{n^2 - 1} \right)}$$

whence,

$$L = \frac{n^2 - 1}{4\pi^2 f_1^2 (C_2 - C_1)}$$

which is the general true inductance formula.

Example : If $f_2 = 600$ kc, $f_1 = 1,200$ kc, $C_2 = 130 \mu\mu\text{f}$, $C_1 = 30 \mu\mu\text{f}$, then

$$n = \frac{1,200}{600} = 2$$

$$L = \frac{2^2 - 1}{4\pi^2(1.2 \times 10^3)^2(130 - 30)10^{-12}}$$

$$= 5.28 \times 10^{-4} \text{ h}$$

$$= 528 \mu\text{h}$$

and

$$C_d = \frac{C_2 - 4C_1}{3}$$

$$= \frac{130 - (4 \times 30)}{3} = 3.33 \mu\mu\text{f}$$

For simplicity of calculation and equipment design, it is common to choose the frequencies as the fundamental and the second harmonic so that

$$\frac{f_1}{f_2} = 2 \quad L = \frac{3}{4\pi^2 f_1^2 (C_2 - C_1)} \quad \text{or} \quad \frac{0.07599}{f_1^2 (C_2 - C_1)}$$

Note also that this gives a method for the measurement of distributed capacitance

from the general formula (5) $C_d = \frac{C_2 - n^2 C_1}{n^2 - 1}$ or for the more usual case where $n = 2$, $C_d = \frac{C_2 - 4C_1}{3}$.

True inductance measurements by this method are sometimes used but are generally spurned in favor of the simpler 1,000-cycle bridge measurements.

True inductance or two-frequency inductor matching is often used in the production testing of inductors to ensure more accurate "tracking" of the r-f part of single-dial radio sets.

7. Two-frequency Test. The test equipment (Fig. 3) consists of two oscillators, one used as a reference and the other used with the inductor under test. The signals

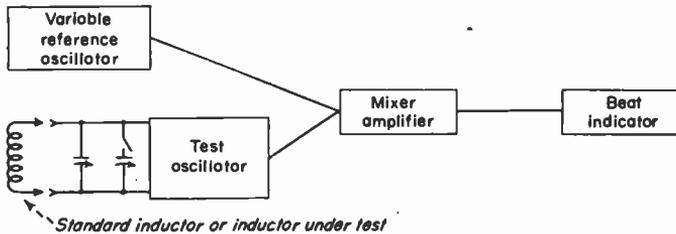


FIG. 3. Two-frequency or true-inductance matching setup.

from the two oscillators are mixed, fed into an amplifier and thence to either an audible or a visible beat indicator to show when the two oscillators are at the same frequency. The reference oscillator is set to the lower of the two frequencies (in the example, 600 kc). A standard inductor, either tested in a radio set or accurately measured as to its true inductance, is placed in a test jig so that it forms the inductance for the second oscillator circuit. A variable capacitor is shunted across the inductor. This capacitor is sufficient to tune the inductor to the second harmonic of the reference oscillator (in the example, 1,200 kc). A switch is provided that allows approximately three times that capacitance to be added in parallel when desired. With the additional capacitance, the inductor will tune to the fundamental frequency (600 kc). The standard inductor is tuned to the second harmonic and then, without disturbing this capacitor, the second capacitor is added and with it the inductor is tuned to the fundamental frequency. Both capacitor readings are noted. An unknown inductor

may now be matched to the standard by removing the standard from the test jig and inserting the other inductor. That inductor is then adjusted until both capacitor readings are within the desired tolerance of the standard. The difference in capacitance readings between the unknown and the standard at the second harmonic gives an indication of the differences in their distributed capacitances, and the difference in readings at the fundamental frequency gives the difference in true inductance between the two.

8. Measurement of Distributed Capacitance. One very common method of measuring distributed capacitance has already been discussed under measurement of inductance. That method must be used with caution since it depends upon the difference of two numbers of much greater magnitude than the resulting distributed capacitance and is, therefore, liable to introduce large errors.

It is possible to determine the natural resonant frequency of an inductor by placing it in parallel with another resonant circuit whose resonant frequency is calibrated and can be varied. The frequency of the calibrated circuit is varied until the introduction of the inductor being measured does not affect the tuning but only the circuit Q . This is the natural resonant frequency of the inductor since at that point it has no reactance but acts as a pure resistance loss. Assuming that the inductance remains relatively constant over a range of frequencies, the distributed capacitance can then be found by measuring the capacitance required to tune the inductor to some lower frequency, far removed from the natural resonant frequency, and then calculating the distributed capacitance by the formula

$$C_d = \frac{f^2 C}{f_n^2}$$

where f = lower frequency

C = capacitance to tune to that frequency

f_n = natural resonant frequency

In practice f_n is about ten times f . This method is more accurate than the previous method, especially for distributed capacitances of less than 10 μmf .

9. Measurement of Q . The figure of merit, Q , is most readily measured on Q meters that have been calibrated to read directly over large frequency ranges. The Q meter is based on the fact that the ratio of the voltage developed across the capacitor in a resonant circuit to the voltage introduced into the circuit is equal to Q . Thus a measured current from an oscillator flows through a resistor of very low resistance in series with the inductor to be measured introducing known voltage into the circuit. A built-in calibrated capacitor completes the series circuit. A voltmeter, calibrated directly in Q , measures the voltage developed across the capacitor. This voltmeter has negligible power consumption.

In resonant circuits containing a well-designed, low-loss capacitor, the resistance of the capacitor is negligible as compared with the resistance of the inductor. The total losses can be considered as occurring in the inductor alone. Under these conditions,

$$Q = \frac{f_r}{(f_2 - f_1)}$$

where f_r is the resonant frequency, f_1 and f_2 are the frequencies above and below resonance at which the power dissipated is one-half the power dissipated at resonance, and the current is 0.707 times the current at resonance. Also

$$Q = \frac{C_r}{C_2 - C_1}$$

where C_r is the capacitance at resonance, and C_1 and C_2 are the capacitances corresponding to frequencies f_1 and f_2 . With these formulas, Q measurements may be made with equipment usually available around a laboratory.

10. Measurement of Mutual Inductance. The simplest method of measuring mutual inductance is the Felici balance in which the unknown mutual inductance is

compared with a variometer whose mutual inductance with respect to rotation has been calibrated. The primaries of both inductors are connected in series across an a-c source, usually 1,000 cycles. The secondaries are also connected in series, so that the voltages induced in them oppose each other, and in series with a null detector, usually a pair of headphones with a suitable amplifier. At the null, the mutual inductance of the unknown is equal to the mutual inductance of the known calibrated variometer.

If an inductance bridge is available, the mutual inductance, M , may be found by measuring the total inductance of the primary and secondary inductances in series aiding and then in series bucking. M is calculated from the formula

$$L_M = \frac{L_{\text{aiding}} - L_{\text{bucking}}}{4}$$

Coefficient of coupling is a term sometimes encountered in working with coupled inductors. This coefficient, K , is the ratio of the mutual inductance to the geometric mean of the primary and secondary inductances, or

$$K = \frac{L_M}{\sqrt{L_p L_s}}$$

TYPES OF INDUCTORS

11. Straight Conductors. Since any conductor carrying a current has the property of self-inductance, the simplest form of inductor is a straight wire. Straight wires are used as inductors in h-f applications where the inductance must be very low. In some cases the inductance is furnished by the current flowing along the plates of the variable tuning capacitor. At the frequencies at which straight conductors become important as inductors, many other complications arise not within the scope of a discussion of inductors; therefore, they are simply mentioned here without any detail.

For reference, a straight piece of No. 10 AWG wire 1 in. long has an inductance of about 0.015 μ h at low frequencies, if the return conductor is assumed to be remote.

12. Single-layer Solenoid Windings. The single-layer solenoid winding is probably the most important type of inductor for frequencies from about 1 to 100 Mc, or higher.

It has the advantage of being fairly simple. Its inductance can be varied by spacing a fraction of the total turns away from the main body of the winding (Fig. 4). Then by moving a few of the spaced turns nearer to the main body, the inductance can be raised slightly. By moving a few turns away from the main body to the spaced portion, the inductance can be lowered slightly. Such adjustment of inductance is important in the tracking of the r-f portions of superheterodyne receivers.



FIG. 4. Solenoid with spaced section for adjustment of inductance.

The single-layer solenoid winding has the further advantage of low distributed capacitance, since the first and last turns which have the greatest voltage developed between them are relatively far apart. The voltage between any two turns is only $1/n$ of the total voltage if n is the total number of turns.

This type of winding can be made to have a fairly constant inductance with temperature changes since it can be wound in grooves in the coil form or firmly cemented to the form.

For a given design, Q variations of ± 5 per cent and inductance variations of ± 2 per cent can be expected in the normal production of solenoids. In the case of coils used in broadcast receivers whether the windings are solenoids, universals, progressive universals, or banks, the common commercial inductance tolerances after adjustment are $\pm 1/2$ per cent for oscillator coils and ± 1 per cent for r-f and antenna coils. See Table 1 for a list of common h-f solenoids.

Table 1. High-frequency Solenoids

OD. in.	Wall. in.	Wire	Tpi*	Turns	Inductance. μ h	Q	Frequency. Mc
1	$\frac{3}{8}$	20E	20	$5\frac{1}{8}$	1.1	105	7
1	$\frac{3}{8}$	20E	20	$5\frac{1}{4}$	0.9	95	8
$\frac{7}{8}$	$\frac{3}{8}$	22E	22	$7\frac{1}{2}$	1.6	165	12
$\frac{3}{4}$	$\frac{3}{8}$	22T	22	$10\frac{1}{2}$	1.7	146	7
$\frac{3}{4}$	$\frac{3}{8}$	22T	22	$7\frac{1}{2}$	1.1	165	14
$\frac{3}{4}$	$\frac{1}{2}$	20E	19	$11\frac{1}{8}$	2.1	127	7
$\frac{3}{4}$	$\frac{1}{2}$	20E	20	$9\frac{1}{2}$	1.2	165	14
$\frac{3}{4}$	$\frac{1}{2}$	22F	16	$12\frac{1}{2}$	2.2	115	10
$\frac{3}{4}$	$\frac{1}{2}$	22T	22	$7\frac{1}{2}$	1.1	165	14
$\frac{3}{16}$	$\frac{3}{8}$	24E	30	15	2.7	145	10
$\frac{3}{16}$	$\frac{3}{8}$	24E	30	$12\frac{1}{4}$	1.9	137	10
$\frac{1}{2}$	$\frac{1}{2}$	22E	34	$19\frac{1}{4}$	3.0	133	12
$\frac{1}{2}$	$\frac{1}{2}$	22E	36	$24\frac{1}{2}$	4.3	112	7

* Tpi is turns per inch for grooving.
 NOTE: All forms are phenolic, grooved.

13. Multilayer Solenoids. Multilayer solenoids are usually wound with the first layer running from left to right, the next running from right to left, etc (Fig. 5). If the layers are separated by paper, the winding is known as a "paper-layer" solenoid. This type of winding has the disadvantage of high distributed capacitance since the beginning of one layer and the end of the next layer, which are at high potential difference, are directly above one another. Multilayer solenoids are used at power and audio frequencies but never at radio frequencies where distributed capacitance is important.

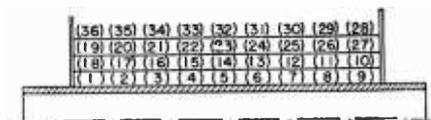


FIG. 5. Cross section of multilayer solenoid winding.

14. Universal Windings. The universal winding is a comparatively narrow multilayer winding in which the conductor spirals back and forth across the width of the winding as the winding is built up. This results in a self-supporting lattice type of construction. By changing the ratio of the pitch of this spiral to the number of turns, varying numbers of crossovers per turn and turns per layer can be obtained. The

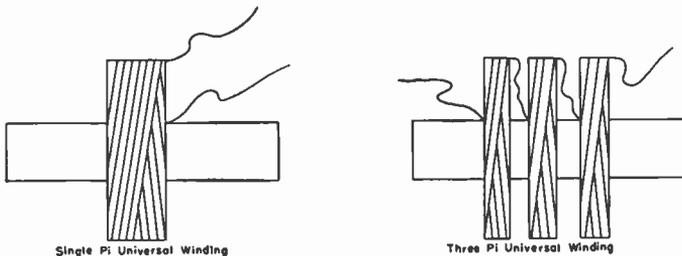


FIG. 6. Single- and three-pi windings.

distributed capacitance is fairly low since turns with high potential difference simply cross each other and do not lie parallel to one another for appreciable distances. Large inductances can be efficiently put into small spaces through the use of universal windings. They are commonly found in r-f chokes, i-f transformers, high-impedance primaries, etc.

By making the universal winding in several sections of pi's side by side on the form, the distributed capacitance can be further reduced and, since the field pattern

Table 2. Typical Universal Wound Choke Coils

All coils are two-cross, single-pi universals. Coils of 0.5 to 20 mh are $\frac{1}{4}$ in. wide wound with No. 36 SSE wire. Coils of 25 to 70 mh are wound with No. 38 SSE wire and are $\frac{5}{16}$ in. wide. All values are approximate.

Inductance, mh	Turns	OD, in.	D-c resistance, ohms	Distributed capacitance, μf
0.5	185	0.60	12	11.3
1.0	260	0.63	17	11.0
1.5	320	0.67	21	10.7
2.0	370	0.69	26	10.2
2.5	400	0.71	29	9.6
3.0	450	0.73	31	9.5
4.0	510	0.76	37	7.6
5.0	570	0.79	42	7.5
10.0	790	0.90	61	7.1
15.0	940	0.98	78	8.2
20.0	1,060	1.04	93	9.8
25.0	1,390	0.83	159	8.8
30.0	1,450	0.85	183	9.0
40.0	1,680	0.90	219	9.5
50.0	1,850	0.97	269	7.4
60.0	2,000	1.01	274	6.9
70.0	2,150	1.07	294	6.6

Table 3. Effect of Form Diameter on Q^*

Form diam, in.	Q in air	Q in $1\frac{3}{4}$ -in.-square Al shield	Q in $1\frac{3}{4}$ -in.-diam, round Al shield
$\frac{3}{8}$	125	118	114
$\frac{1}{2}$	132	122	117
$\frac{5}{8}$	131	120	116

* 1-mh coil, 7/41 SSE wire, $\frac{1}{4}$ in. wide, $2\frac{1}{2}$ crosses per turn, single pi.

is elongated, the effect of placing a shield around the winding is diminished. The effective diameter is also decreased, a fact that can be important when powdered-iron cores are used. Honeycomb coils were a type of universal winding with relatively few, widely spaced turns per layer giving a typical "honeycomb" appearance. The universal winding can be varied in inductance somewhat for tracking purposes. It can be reduced by squeezing, which shortens the effective diameter. The inductance may be raised by squeezing the thickness of the winding to increase its effective diameter. Such distortion of the winding is not too desirable since it makes for poor stability, reduces Q , and increases distributed capacitance. Pi windings may be moved closer together or farther apart to accomplish the same end with the same disadvantages.

Table 4. Effect of Width of Winding on Q^*

Width, in.	Q in air
$\frac{3}{16}$	101
$\frac{1}{8}$	105
$\frac{1}{4}$	109

* 1-mh coil, 7/41 SSE wire, 2 crosses per turn, single pi.

Universal windings are commonly used for inductors from $10 \mu\text{h}$ to 1 h. They may have Q 's up to about 150. Common Q variations found in commercial windings of the same design and inductance are ± 10 per cent. The usual inductance tolerance for unadjusted windings is ± 5 per cent. See Tables 2 to 7 for data on universal windings.

Table 5. Effect of Number of Crosses per Turn on Q*

No. crosses	Q in air	Q in 1 3/4-in. square Al shield	Q in 1 3/4-in.-diam round Al shield
1 1/4	120	113	110
2	109	102	99
2 1/5	125	118	114

* 1-mh coil, 7/41 SSE wire, 1/4 in. wide, 3/8-in.-diam form, single pi.

Table 6. Effect of Number of Pi's on Q*

No. pi's	Q in air	Q in 1 3/4-in.-square Al shield	Q in 1 3/4-in.-diam round Al shield
2	131	124	120
3	134	128	123
4	132	126	123

* 1-mh coil, 7/41 SSE wire, pi's 3/2 in. wide spaced 1/8 in., 1/2 in. diam, form 4 crosses per turn

Table 7. Effect of Wire Size on Q;* Relative Cost and Relative OD

Wire size	Q in air	Q in 1 3/4-in.-square Al shield	Q in 1 3/4-in.-diam round Al shield	Relative cost	Relative OD
3/40	103	97	95	1 (lowest)	1 (smallest)
5/40	119	108	105	5	5
5/41	121	113	110	3	2
9/42	130	120	116	2	1 (smallest)
7/41	133	120	114	4	3
19/41	139	122	107	7	7 (largest)
10/42	140	128	123	6	6
1 3/43	157	139	133	8	4
1 5/44	176	149	142	9 (highest)	5

* 1-mh coil, 3 3/5 crosses per turn, 3/16 in. wide, 1/2-in.-diam form, single pi.

An article by A. W. Simon* covers the general theory of universal windings and gives other references to work done on the mathematical study of these windings.

15. Bank Windings. Banked or bank windings are another form of multilayer windings in which a special arrangement of the turns in the various layers tends to reduce distributed capacitance. The arrangement is shown in Fig. 7. This type of winding is difficult to produce and has, therefore, lost favor.

Commercial Q variation for a given design is ±10 per cent while the inductance variation is about ±5 per cent for unadjusted windings. See Table 8 for a comparison of bank and progressive universal windings.

16. Progressive Universal Windings. To a large degree, progressive universal windings have supplanted the bank winding. This winding is produced in a manner similar to the universal winding except that instead of spiraling back and forth in the same area, the spiraling also progresses along the form and gives a winding that is spread out horizontally rather than built up vertically. A

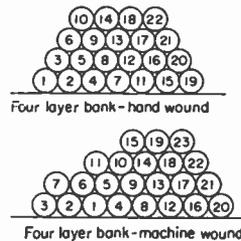


FIG. 7. Cross section of bank windings.

* Proc. IRE, January, 1945.

simple progressive winding looks somewhat like a bank winding. Because it is possible

to vary the crossovers per turn as well as the rate of progression, a great variety of windings is possible. This type of winding helps make possible the high- Q permeability-tuned inductors to be discussed later.

The commercial variations in Q and inductance of a given design of a progressive universal winding are the same as those for universal and bank windings, ± 10 and ± 5 per cent, respectively.

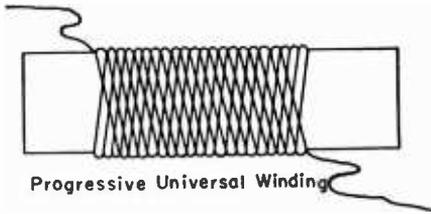


FIG. 8. Appearance of progressive universal winding.

An article by A. W. Simon* gives some of the theory of this complex winding type.

Table 8. Comparison of Progressive Universal and Two-layer Bank Windings

Air inductance of all windings 250 μ h; wire 7/41 SSE

Form diam, in	600 kc		1,000 kc		1,400 kc	
	Prog.	Bank	Prog.	Bank	Prog.	Bank
Q in air						
$\frac{5}{8}$	110	120	142	148	148	147
$\frac{3}{4}$	123	134	160	166	165	165
$\frac{7}{8}$	131	136	170	174	171	176
1	134	143	172	180	169	177
Q in $1\frac{1}{8}$ - by $1\frac{3}{8}$ -in. Al shield						
$\frac{5}{8}$	82	106	110	135	119	141
$\frac{3}{4}$	80	106	121	137	130	143
$\frac{7}{8}$	79	101	120	130	130	137
1	84	99	113	128	122	135

17. Other Types of Windings. There are a few other types of windings of more or less importance. For example, the flat spiral is widely used in loop antennas for broadcast receivers. Spider-web, basket-weave, and similar windings are cumbersome and difficult to produce. The toroidal winding has its conductor wound on the surface of a toroid. When carefully made, the field is almost completely self-contained. It requires little shielding and is sometimes used in making inductance standards.

18. Iron-cored Inductors. Since the permeability of the medium has a definite effect on the flux that causes self-inductance, it is obvious that the inductance of an air-cored winding can be increased by introducing material, such as iron, into the winding to increase the permeability of the medium. However, the introduction of the material also introduces a new source of losses. Magnetic losses due to hysteresis and eddy currents were discussed in a preceding section. At low frequencies, in the power or audio region, laminating the iron will substantially reduce the eddy-current losses. As long as large amounts of d.c. are not present in the inductor, which would tend to produce saturation of the iron and hence a nonlinear inductance as the current is varied, the only major problem is to get as much permeability as possible in order to get as much inductance as possible in a small space.

At radio frequencies it is not possible or practical to make laminations thin enough to reduce the losses to a sufficient degree. An entirely different technique is, therefore, employed in making the core: powder metallurgy. In this process the iron is in a very

* *Proc. IRE*, December, 1945.

finely powdered state, produced either mechanically or chemically. The powder is combined with suitable binders, compressed, and sintered to produce a very compact mass which nonetheless consists of separate particles with an average size of 10 microns or less.

Table 9. Toroidal Iron-core Coils

This table* gives data derived from measurements on toroidal cores of the most important iron powders available on the market. All toroids were prepared in the same way for reasons of comparison, i.e., with 4 per cent bakelite (type 7095), 2 per cent Sterotex for lubrication, 50 tons per sq in. pressure, and 50 g weight. Their OD is 2.25 in., ID 1.50 in., and they are wound with a single layer of 200 turns of No. 25 magnet wire.

The data were obtained for the most part according to the method of V. E. Legg, *Bell System Tech. J.* 16, 39, 1936. Thus, the following units apply:

- Density, g/cm³
- Permeability, gausses gilbert/cm
- Hysteresis loss coefficient, 10⁻³ ohm/h/gauss cycle
- Eddy-current loss coefficient, 10⁻⁶ ohm/h/cycle squared
- Magnetic stability, percentage change of permeability after 4 amp passage through winding

Material	Density	Permeability	Hysteresis	Eddy current coeff	Stability
Carbonyl L.....	5.7	24.8	3.1	0.13	+1.4
Carbonyl C.....	5.7	16.7	1.1	0.14	+0.6
Carbonyl E.....	4.9	10.4	0.3	0.11	+0.6
Carbonyl TH.....	4.9	9.6	0.3	0.10	+0.18
Carbonyl SF.....	4.8	8.1	0.3	0.10	+0.15
Electrolytic.....	5.6	23.4	2.4	0.33	-0.17
Hydrogen reduced, 1.....	5.6	18.4	2.6	0.12	+1.3
Hydrogen reduced, 2.....	5.5	16.9	1.0	0.12	+1.6
Hydrogen reduced, 3.....	5.2	12.5	3.1	0.11	+1.5
Magnetite, 1.....	4.1	7.9	9.1	11.5	-2.1
Magnetite, 2.....	4.0	5.7	6.8	0.21	-1.6
Magnetite, 3.....	3.5	3.1	0.3	0.085	+0.10

* All data from General Aniline and Film Corporation.

Carbonyl iron powders are obtained by thermal decomposition of iron pentacarbonyl. There are five different grades in production, designated as L, C, E, TH, and SF powder. Each of these is obtained by special process methods and has its special field of application. The particles making up the powders E, TH, and SF are spherical with a characteristic structure of increasingly larger shells. The particles of L and C are made up of homogeneous spheres and agglomerates.

The chemical analysis, the weight-average particle size as determined in a Roller air analyzer, the "tap density," i.e., the density of the loose powder packed by tapping the container in a prescribed manner, and the apparent density or bulking factor as determined in a Scott volumeter are given in Table 10 for the five grades.

Table 10*

Grade	Chemical analysis, per cent			Wt—ave. diam, microns	Tap density, g/cm ³	Apparent density, g/cm ³
	Carbon	Oxygen	Nitrogen			
L	0.005-0.03	0.1 -0.2	0.005-0.05	20	3.5-4.0	1.8-3.0
C	0.03 -0.12	0.1 -0.3	0.01 -0.1	10	4.4-4.7	2.5-3.0
E	0.65 -0.80	0.45-0.60	0.6 -0.7	8	4.4-4.7	2.5-3.5
TH	0.5 -0.6	0.5 -0.7	0.5 -0.6	5	4.4-4.7	2.5-3.5
SF	0.5 -0.6	0.7 -0.8	0.5 -0.6	3	4.7-4.8	2.5-3.5

* All data from General Aniline and Film Corporation.

With reference to the chemical analysis shown in Table 10, it should be noted that spectroscopic analysis shows the rest to be iron with other elements present in traces only.

Carbonyl iron powders are primarily useful as electromagnetic material over the entire communication frequency spectrum.

Table 11 gives relative Q values and effective permeabilities for the different grades of carbonyl iron powder. The values are derived from measurements on straight cylindrical cores placed in simple solenoidal coils. Although the data were not obtained at optimum conditions, the Q values as expressed in percentage of the best core give an indication of the useful frequency ranges for the different powder grades.

Table 11*

Grade	Effective permeability at 1 kc	Relative quality factor at				
		10 kc	150 kc	200 kc	1 Mc	100 Mc
L	4.16	100	96	90	43	1
C	3.65	94	100	98	72	3
E	3.09	81	94	100	97	30
TH	2.97	81	93	98	100	54
SF	2.17	62	71	78	84	100

* All data from General Aniline and Film Corporation.

Note: Q values actually measured can be obtained by multiplying the rows, respectively, by 0.78, 1.09, 1.25, 2.63, and 1.62.

L and C powders are also used as powder metallurgical material because of their low sintering temperatures, high tensile strengths, and other very desirable qualities. Sintering begins below 500°C, and tensile strengths reach 150,000 psi.

19. Variable Inductors. Probably the simplest form of variable inductor is a tapped winding in which the inductance is varied by taking more or less of the total available winding. This is a step-by-step method of variation and usually not satisfactory. Another form, which is rather bulky, difficult to manufacture, and now little used, is the variometer. This consists of two similar windings so disposed that they are closely coupled and can be changed from a mutual aiding to mutual bucking position in a continuous manner. This is accomplished by placing one winding on the inner surface of a sphere and the other on the outer surface of a slightly smaller sphere within it. The two windings are connected in series. Rotating one sphere with respect to the other gives the desired change in mutual and, hence, in total inductance.

Much more common is the variation of inductance through the variation of the permeability of the core. This can be accomplished by the use of either a movable powdered-iron core or a copper core. The former increases the permeability while the latter decreases it because of the introduction of eddy currents in the core. The eddy currents always buck the inducing flux and, hence, decrease the self-inductance.

The winding, commonly a solenoid, universal or progressive universal, is so arranged that an iron or copper core can be moved into its center; thus, the air core is replaced by a core of positive or negative permeability which correspondingly raises or lowers the total inductance of the winding. Inserting a powdered-iron core will raise the inductance as much as fourteen times; inserting a copper core will lower the inductance to 1/1.4 of its air-core value.

Powdered-iron cores are used at all frequencies up to 100 Mc, or slightly higher. Within the frequency range for which they are designed, iron cores can be obtained with effective permeabilities up to 14. They can also effect considerable increase in Q despite the fact that the cores themselves introduce some losses.

No good means of specifying the permeabilities of powdered-iron cores has yet been devised. Some work has been done on absolute permeabilities through the use

of toroid windings. Although this showed promise, the method was far too cumbersome for practical use. The present practice is to choose the core for the particular application.

The effective permeability of the core for the particular winding is simply the ratio of the inductance with the core in the winding to the inductance with an air core. The effective permeability is definitely affected by the winding design, as will be discussed later.

Since the core composition that gives the greatest permeability is not always the one that gives the best Q , the choice of core is a compromise.

By properly designing a tapered winding, the frequency to which it is tuned by introducing the iron core into the winding can be made linear with linear motion of the core. Furthermore, by the use of such variable-pitch windings, both solenoid and progressive universal types, and by use of movable iron cores, an r-f tuning unit of the two- or three-gang type may be made using permeability-tuned inductors rather than the usual gang capacitors. Such permeability-tuned units can be made small, light, and highly efficient. "Tracking" is accomplished by matched windings and iron cores.

In general, a circuit that contains a permeability-tuned inductor with a stable silvered-mica type of capacitor has far greater frequency stability with time and temperature than a circuit containing a fixed inductor and a variable mica capacitor.

Permeability tuning may also be accomplished by introducing into the winding a slug of low-resistance conducting material. The eddy currents induced in the slug buck the flux of the inductor to produce apparent negative permeabilities as high as 1.4.

Copper cores are used at frequencies above about 40 Mc. At such frequencies Q 's are relatively high, and the decreasing Q caused by the core is not serious. Use of copper cores allows a maximum conductor length to be employed since the core decreases the inductance. This is mechanically desirable in h-f winding design.

Figure 9 shows the over-all selectivity characteristics of two two-stage 455-kc i-f systems. Curve 1 is that of an air-trimmed, four-pi universal-wound coil system. Curve 2 is that of a single-pi, universal-wound, permeability-tuned coil system. The tubes used in the i-f amplifiers were 6A8, 6K7, 6K7, 6Q7. Despite the fact that less expensive single-pi coils are used, the increased sharpness of resonance due to the use of iron-cored, permeability-tuned units is obvious.

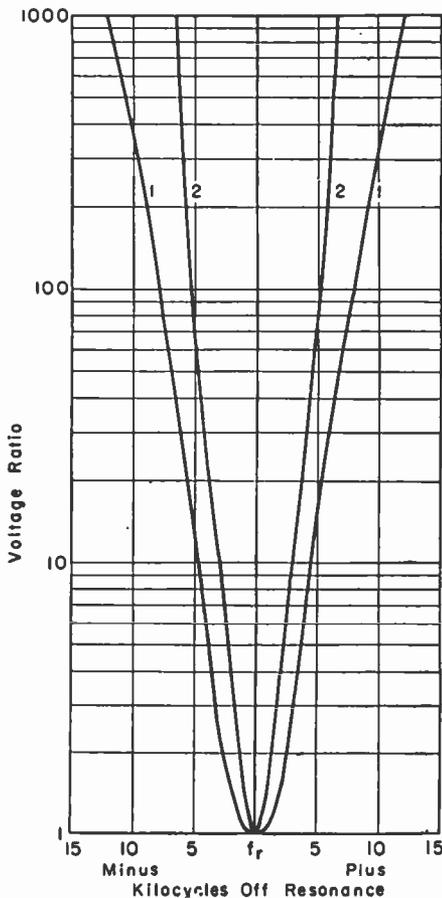


Fig. 9. Selectivity characteristics of two-stage 455-kc i-f amplifiers using 6A8, 6K7, 6K7, and 6Q7 tubes. Curve 1: air-trimmed, four-pi universal air-core units; curve 2: single-pi universal permeability-tuned units.

20. Inductor Design. The design of inductors is, for the most part, somewhat more of an art than a science. Many practical factors influence it to such an extent that the theoretically best inductors are not obtainable. Available space, shielding requirements, materials of construction, and cost are usually the controlling factors. Where they are practical, the following formulas for the calculation of inductance are given, but for the most part some examples of the effects of the various variables give a basis for future design work.

The design is somewhat influenced by the end use. However, the principles are the same. It is necessary to stress only the particular characteristic desired in the original design.

Since most inductors now being designed and used in the radio engineering field are for low power applications, as in receiving equipment, the practical data given in this section pertain to such applications. In high-powered transmitting units less thought is given to space and cost, and more thought is given to losses. Most such inductors are designed for single-frequency operation. They may operate at such high power that cooling is necessary. Low-power-transmitting inductors are similar to receiving inductors.

CALCULATION OF INDUCTANCE

It is possible to calculate the true (l-f) inductance of some types of air-cored coils with a fair degree of accuracy. *National Bureau of Standards Circular 74* has long been the most authoritative source of such information. Some of the more useful formulas contained therein are reproduced here.¹ It must be remembered that for r-f inductors the actual apparent inductance differs from the calculated true inductance because of distributed capacitance. For that reason, inductance calculations are generally used only for a starting point in the final design.

In the following formulas all dimensions are expressed in centimeters, and the inductance is in microhenrys.

21. Straight Round Wire. If l is the length of the wire, d is the diameter of the cross section, and μ is the permeability of the material of the wire,

$$L_0 = 0.002l \left(\log_e \frac{4l}{d} - 1 + \frac{\mu}{4} \right) \quad (6)$$

$$= 0.002l \left(2.303 \log_{10} \frac{4l}{d} - 1 + \frac{\mu}{4} \right) \quad (7)$$

If $\mu = 1$ (for all materials except iron),

$$L_0 = 0.002l \left(2.303 \log_{10} \frac{4l}{d} - 0.75 \right) \quad (8)$$

The return conductor is assumed to be remote. These formulas give the l-f inductance.

As the frequency increases, the inductance decreases, its value at infinite frequency being

$$L_\infty = 0.002l \left(2.303 \log_{10} \frac{4l}{d} - 1 \right) \quad (9)$$

A general expression for the inductance at any frequency is

$$L = 0.002l \left(2.303 \log_{10} \frac{4l}{d} - 1 + \mu\delta \right) \quad (10)$$

¹ Ross, E. B., and F. W. Grover, *Nat. Bur. Standards Sci. Paper* 169; Grover, F. W., *Nat. Bur. Standards Sci. Papers* 320, 1917; 455, 1922; 468, 1923. See for coil design and calculation, especially at low frequencies, Morgan Brooks and H. M. Turner, *Inductance of Coils, Bull* 53, *Univ. Illinois Eng. Exp. Sta.*, Jan. 8, 1912. Grover, F. W., "Inductance Calculations," Van Nostrand, 1946.

The quantity δ is obtained from the table below, as a function of the argument x , where

$$x = 0.1405d \sqrt{\frac{\mu f}{\rho}} \tag{11}$$

and f is the frequency and ρ is the volume resistivity of the wire in microhm-centi-

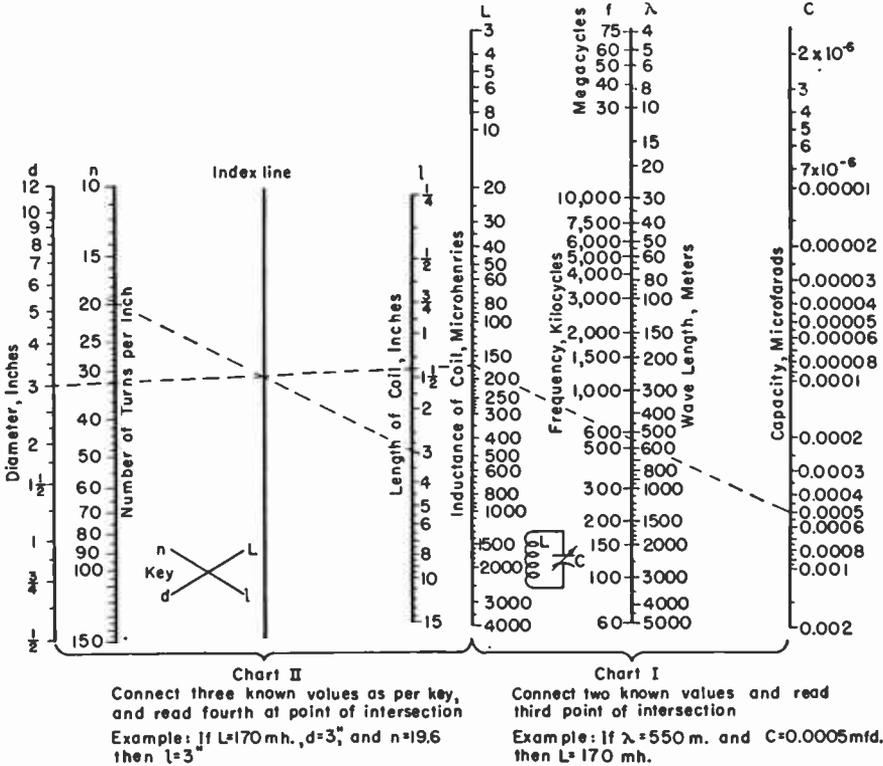


FIG. 10. Inductance-design chart.

meters. For copper at 20°C,

$$x_c = 0.1071d \sqrt{f}$$

This quantity δ will be used in several of the following formulas without further definition.

Value of δ in Inductance Formulas

x	δ	x	δ	x	δ	x	δ	x	δ	x	δ
0	0.250	2.5	0.228	6.0	0.116	12.0	0.059	25.0	0.028	70.0	0.010
0.5	0.250	3.0	0.211	7.0	0.100	14.0	0.050	30.0	0.024	80.0	0.009
1.0	0.249	3.5	0.191	8.0	0.088	16.0	0.044	40.0	0.0175	90.0	0.008
1.5	0.247	4.0	0.1715	9.0	0.078	18.0	0.039	50.0	0.014	100.0	0.007
2.0	0.240	5.0	0.139	10.0	0.070	20.0	0.035	60.0	0.012	∞	0.000

22. Two Parallel Round Wires—Return Circuit. The current is assumed to flow in opposite directions in two parallel wires of length l and diameter d , the distance between centers of wires being D . Then

$$L = 0.004l \left(2.303 \log_{10} \frac{2D}{d} - \frac{D}{l} + \mu\delta \right) \quad (12)$$

This neglects the inductance of the wires connecting the two main wires. If these wires are long, their inductance may be calculated by Eq. (10) and added to the result from Eq. (12), or the whole system may be treated as a rectangle and the inductance calculated by Eq. (14).

23. Square of Round Wire. The length of one side of the square is denoted by a ; other letters have already been defined.

$$L = 0.008a \left(2.303 \log_{10} \frac{2a}{d} + \frac{d}{2a} - 0.774 + \mu\delta \right) \quad (13)$$

Value of K in Eq. (17)

Diam to length	K	Difference	Diam to length	K	Difference	Diam to length	K	Difference
0.00	1.0000	-0.0209	2.00	0.5255	-0.0118	7.00	0.2584	-0.0047
.05	.9791	203	2.10	.5137	112	7.20	.2537	45
.10	.9588	197	2.20	.5025	107	7.40	.2491	43
.15	.9391	190	2.30	.4918	102	7.60	.2448	42
.20	.9201	185	2.40	.4816	97	7.80	.2406	40
0.25	0.9016	-0.0178	2.50	0.4719	-0.0093	8.00	0.2366	-0.0094
.30	.8838	173	2.60	.4626	89	8.50	.2272	86
.35	.8665	167	2.70	.4537	85	9.00	.2185	79
.40	.8499	162	2.80	.4452	82	9.50	.2106	73
.45	.8337	156	2.90	.4370	78	10.00	.2033	
0.50	0.8181	-0.0150	3.00	0.4292	-0.0075	10.0	0.2033	-0.0133
.55	.8031	146	3.10	.4217	72	11.0	.1903	113
.60	.7885	140	3.20	.4145	70	12.0	.1790	98
.65	.7745	136	3.30	.4075	67	13.0	.1692	87
.70	.7609	131	3.40	.4008	64	14.0	.1605	78
0.75	0.7478	-0.0127	3.50	0.3944	-0.0062	15.0	0.1527	-0.0070
.80	.7351	123	3.60	.3882	60	16.0	.1457	63
.85	.7228	118	3.70	.3822	58	17.0	.1394	58
.90	.7110	115	3.80	.3764	56	18.0	.1336	52
.95	.6995	111	3.90	.3708	54	19.0	.1284	48
1.00	0.6884	-0.0107	4.00	0.3654	-0.0052	20.0	0.1236	-0.0085
1.05	.6777	104	4.10	.3602	51	22.0	.1151	73
1.10	.6673	100	4.20	.3551	49	24.0	.1078	63
1.15	.6573	98	4.30	.3502	47	26.0	.1015	56
1.20	.6475	94	4.40	.3455	46	28.0	.0959	49
1.25	0.6381	-0.0091	4.50	0.3409	-0.0045	30.0	0.0910	-0.0102
1.30	.6290	89	4.60	.3364	43	35.0	.0808	80
1.35	.6201	86	4.70	.3321	42	40.0	.0728	64
1.40	.6115	84	4.80	.3279	41	45.0	.0664	53
1.45	.6031	81	4.90	.3238	40	50.0	.0611	43
1.50	0.5950	-0.0079	5.00	0.3198	-0.0076	60.0	0.0528	-0.0061
1.55	.5871	76	5.20	.3122	72	70.0	.0467	48
1.60	.5795	74	5.40	.3050	69	80.0	.0419	38
1.65	.5721	72	5.60	.2981	65	90.0	.0381	31
1.70	.5649	70	5.80	.2916	62	100.0	.0350	
1.75	0.5579	-0.0068	6.00	0.2854	-0.0059			
1.80	.5511	67	6.20	.2795	56			
1.85	.5444	65	6.40	.2739	54			
1.90	.5379	63	6.60	.2685	52			
1.95	.5316	61	6.80	.2633	49			

24. Rectangle of Round Wire. The sides of the rectangle are a and a_1 and the diagonal $g = \sqrt{a^2 + a_1^2}$. Then

$$L = 0.00921 \left[(a + a_1) \log_{10} \frac{4aa_1}{d} - a \log_{10} (a + g) - a_1 \log_{10} (a_1 + g) \right] + 0.004 \left[\mu\delta(a + a_1) + 2 \left(g + \frac{d}{2} \right) - 2(a + a_1) \right] \quad (14)$$

25. Grounded Horizontal Wire. The wire is assumed to be parallel to the earth which acts as the return circuit. In addition to symbols already used, h denotes the height of the wire above ground. Then

$$L = 0.004605l \left[\log_{10} \frac{4h}{d} + \log_{10} \left(\frac{l + \sqrt{l^2 + \frac{d^2}{4}}}{l + \sqrt{l^2 + 4h^2}} \right) \right] + 0.002 \left(\sqrt{l^2 + 4h^2} - \sqrt{l^2 + \frac{d^2}{4}} + \mu l \delta - 2h + \frac{d}{2} \right) \quad (15)$$

26. Circular Ring of Circular Section. If a is the mean radius of the ring,

$$L = 0.01257a \left(2.303 \log_{11} \frac{16a}{d} - 2 + \mu\delta \right) \quad (16)$$

provided that $d/2a \leq 0.2$.

27. Single-layer Coil or Solenoid.

$$L = \frac{0.0395a^2n^2}{b} K \quad (17)$$

where n is the number of turns, a is the radius of the coil measured from the axis to the center of the wire, b is the length of the coil, and K is a function of $2a/b$, the value of which may be determined by means of the table shown on page 134.

28. Multilayer Coils: Circular Coils of Rectangular Cross Section. For long coils of a few layers, the following formula may be used:

$$L = L_s - \frac{0.0126n^2ac}{b} (0.693 + B_s) \quad (18)$$

where L_s is the inductance calculated by Eq. (17), n and b are the same as in Eq. (17), a is the radius of coil measured from axis to center of winding cross section, c is the radial depth of winding, and B_s is the correction given below.

Value of B_s in Eq. (18)

b/c	B_s										
1	0.0000	6	0.2446	11	0.2844	16	0.3017	21	0.3116	26	0.3180
2	0.1202	7	0.2563	12	0.2888	17	0.3041	22	0.3131	27	0.3190
3	0.1753	8	0.2656	13	0.2927	18	0.3062	23	0.3145	28	0.3200
4	0.2076	9	0.2730	14	0.2961	19	0.3082	24	0.3157	29	0.3209
5	0.2292	10	0.2792	15	0.2991	20	0.3099	25	0.3169	30	0.3218

For short multilayer coils, the dimensions shown in Fig. 11 are used. Two formulas are required, one for use when $b > c$, and the other for use when $b < c$. In the first case,

$$L = 0.01257an^2 \left[\left(1 + \frac{b^2}{32a^2} + \frac{c^2}{96a^2} \right) \log_e \frac{8a}{d} - y_1 + \frac{b^2}{16a^2} y^2 \right] = 0.01257an^2 \left[2.303 \left(1 + \frac{b^2}{32a^2} + \frac{c^2}{96a^2} \right) \log_{10} \frac{8a}{d} - y_1 + \frac{b^2}{16a^2} y^2 \right] \quad (19)$$

When $b < c$,

$$L = 0.01257an^2 \left[\left(1 + \frac{b^2}{32a^2} + \frac{c^2}{96a^2} \right) \log_e \frac{8a}{d} - y_1 + \frac{c^2}{16a^2} y^3 \right] \\ = 0.01257an^2 \left[2.303 \left(1 + \frac{b^2}{32a^2} + \frac{c^2}{96a^2} \right) \log_{10} \frac{8a}{d} - y_1 + \frac{c^2}{16a^2} y^3 \right] \quad (20)$$

y_1 , y_2 , and y_3 may be obtained from the table shown below. These formulas are quite

Value of Constants in Eqs. (19) and (20)

b/c or c/b	y_1	c/b	y_2	b/c	y_3
0	0.5000	0	0.125	0	0.597
0.025	0.5253				
0.05	0.5490	0.05	0.127	0.05	0.599
0.10	0.5924	0.10	0.132	0.10	0.602
0.15	0.6310	0.15	0.142	0.15	0.608
0.20	0.6652	0.20	0.155	0.20	0.615
0.25	0.6953	0.25	0.171	0.25	0.624
0.30	0.7217	0.30	0.192	0.30	0.633
0.35	0.7447	0.35	0.215	0.35	0.643
0.40	0.7645	0.40	0.242	0.40	0.654
0.45	0.7816	0.45	0.273	0.45	0.665
0.50	0.7960	0.50	0.307	0.50	0.677
0.55	0.8081	0.55	0.344	0.55	0.690
0.60	0.8182	0.60	0.384	0.60	0.702
0.65	0.8265	0.65	0.427	0.65	0.715
0.70	0.8331	0.70	0.474	0.70	0.729
0.75	0.8383	0.75	0.523	0.75	0.742
0.80	0.8422	0.80	0.576	0.80	0.756
0.85	0.8451	0.85	0.632	0.85	0.771
0.90	0.8470	0.90	0.690	0.90	0.786
0.95	0.8480	0.95	0.752	0.95	0.801
1.00	0.8483	1.00	0.816	1.00	0.816

accurate as long as the diagonal of the cross section (d Fig. 11) does not exceed the mean radius. The accuracy decreases considerably as b becomes large in comparison with a .

For very accurate results, a correction must be added if the insulation of the wire occupies a considerable percentage of the winding space. This correction is given by

$$\Delta L = 0.01257an \left(2.303 \log_{10} \frac{D}{d} + 0.155 \right) \quad (21)$$

where D is the distance between the centers of adjacent wires, and d is the diameter of the bare wire.

29. Multilayer Square Coil. If n is the number of turns and a is the side of the square measured to the center of the rectangular cross section which has length b and depth c , then

$$L = 0.008an^2 \left(2.303 \log_{10} \frac{a}{b+c} + 0.2235 \frac{b+c}{a} + 0.726 \right) \quad (22)$$

FIG. 11. Multi-layer coil.

If the cross section is square ($b = c$), this becomes

$$L = 0.008an^2 \left(2.303 \log_{10} \frac{a}{b} + 0.447 \frac{b}{a} + 0.033 \right) \quad (23)$$

Equation (21) may be used to correct for insulation by replacing the factor 0.01257 by 0.008. For a single-layer square coil,

$$L = 0.008an^2 \left(2.303 \log_{10} \frac{a}{b} + 0.2231 \frac{b}{a} + 0.726 \right) - 0.008an(A + B) \quad (24)$$

A and *B* are given below, where *d* is the diameter of the bare wire and *D* is the distance between turns, measured to the centers of the wires.

Value of *A* in Eq. (24)

<i>d/D</i>	<i>A</i>	<i>d/D</i>	<i>A</i>	<i>d/D</i>	<i>A</i>
1.00	0.557	0.40	-0.359	0.15	-1.340
0.95	0.506	0.38	-0.411	0.14	-1.409
0.90	0.452	0.36	-0.465	0.13	-1.483
0.85	0.394	0.34	-0.522	0.12	-1.563
0.80	0.334	0.32	-0.583	0.11	-1.650
0.75	0.269	0.30	-0.647	0.10	-1.746
0.70	0.200	0.28	-0.716	0.09	-1.851
0.65	0.126	0.26	-0.790	0.08	-1.969
0.60	0.046	0.24	-0.870	0.07	-2.102
0.55	-0.041	0.22	-0.957	0.06	-2.256
0.50	-0.136	0.20	-1.053	0.05	-2.439
0.48	-0.177	0.19	-1.104	0.04	-2.662
0.46	-0.220	0.18	-1.158	0.03	-2.950
0.44	-0.264	0.17	-1.215	0.02	-3.355
0.42	-0.311	0.16	-1.276	0.01	-4.048

Value of *B* in Eq. (24)

Number of turns, <i>n</i>	<i>B</i>	Number of turns, <i>n</i>	<i>B</i>
1	0.000	40	0.315
2	0.114	45	0.317
3	0.166	50	0.319
4	0.197	60	0.322
5	0.218	70	0.324
6	0.233	80	0.326
7	0.244	90	0.327
8	0.253	100	0.328
9	0.260	150	0.331
10	0.266	200	0.333
15	0.286	300	0.334
20	0.297	400	0.335
25	0.304	500	0.336
30	0.308	700	0.336
35	0.312	1.000	0.336

30. Calculation of Mutual Inductance.¹ The mutual inductance of two parallel coaxial circles may be calculated by the following method: first, calculate

$$\frac{r_2}{r_1} = \sqrt{\frac{\left(1 - \frac{a}{A}\right)^2 + \frac{D^2}{A^2}}{\left(1 + \frac{a}{A}\right)^2 + \frac{D^2}{A^2}}} \quad (25)$$

where *a* is the radius of the smaller circle, *A* the radius of the larger circle, and *D* the

¹ Ross, E. B., and F. W. Grover. *Nat. Bur. Standards Sci. Paper* 169; Grover, F. W., *Nat. Bur. Standards Sci. Papers* 320 and 498.

Values of F in Eq. (26)

r_2/r_1	F	Difference	r_2/r_1	F	Difference	r_2/r_1	F	Difference
0	∞							
0.010	0.05016	-0.00120	0.30	0.008844	-0.000341	0.80	0.0007345	-0.0000604
.011	4897	109	.31	8503	328	.81	6741	579
.012	4787	100	.32	8175	314	.82	6162	555
			.33	7861	302	.83	5607	531
0.013	4687	-0.00093	.34	7559	290	.84	5076	507
.014	4594	87						
.015	4507	81	0.35	0.007260	-0.000280	0.85	0.0004569	-0.0000484
.016	4426	148	.36	6989	270	.86	4085	460
.018	4278	132	.37	6720	260	.87	3625	437
			.38	6460	249	.88	3188	413
0.020	0.04146	-0.00119	.39	6211	241	.89	2775	389
.022	4027	109						
.024	3918	100	0.40	0.005970	-0.000232	0.90	0.0002386	-0.0000365
.026	3818	93	.41	5738	225	.91	2021	341
.028	3725	86	.42	5514	217	.92	1680	316
			.43	5297	210	.93	1364	290
0.030	3639	-0.00081	.44	5087	202	.94	1074	263
.032	3558	76						
.034	3482	71	0.45	0.004885	-0.000195	0.95	0.00008107	-0.00002351
.036	3411	68	.46	4690	189	.96	5756	2046
.038	3343	64	.47	4501	183	.97	3710	1706
			.48	4318	178	.98	2004	1301
0.040	0.03279	-0.00061	.49	4140	171	.99	703	703
.042	3718	58				1.00	0	
.044	3160	55	0.50	0.003969	-0.000166			
.046	3105	53	.51	3803	160	0.950	0.00008170	0.00000494
.048	3052	51	.52	3643	156	.952	7613	482
			.53	3487	150	.954	7131	470
0.050	0.03001	-0.00226	.54	3337	146	.956	6661	458
.060	2775	191				.958	6202	446
.070	2584	164	0.55	0.003191	-0.000141			
.080	2420	144	.56	3050	137	0.960	0.00005756	-0.00000436
.090	2276	128	.57	2913	133	.962	5320	421
			.58	2780	128	.964	4899	409
0.100	0.02148	-0.00116	.59	2652	125	.966	4490	397
.11	2032	104				.968	4093	383
.12	1928	96	0.60	0.002527	-0.000120			
.13	1832	89	.61	2407	117	0.970	0.00003710	-0.00000370
.14	1743	82	.62	2290	113	.972	3340	356
			.63	2177	109	.974	2984	341
0.15	0.01661	-0.00075	.64	2068	106	.976	2643	327
.16	1586	71				.978	2316	312
.17	1515	66	0.65	0.001962	-0.000103			
.18	1449	62	.66	1859	99	0.980	0.00002004	-0.00000296
.19	1387	59	.67	1760	96	.982	1708	278
			.68	1664	93	.984	1430	262
0.20	0.01328	-0.00055	.69	1571	90	.986	1168	242
.21	1273	52				.988	926	223
.22	1221	50	0.70	0.001481	-0.000087			
.23	1171	47	.71	1394	84	0.990	0.00000703	-0.00000201
.24	1124	45	.72	1310	81	.992	502	177
			.73	1228	78	.994	326	148
0.25	0.010792	-0.000425	.74	1150	76	.996	177	115
.26	10366	408				.998	62	62
.27	9958	388	0.75	0.0010741	-0.0000731			
.28	9570	371	.76	10010	704			
.29	9199	355	.77	9306	680			
			.78	8626	653			
			.79	7973	628			

distance between the planes of the two circles. From the table on page 138 the value of F corresponding to the calculated value of r_2/r_1 is obtained. Then

$$M = F \sqrt{Aa} \quad (26)$$

The units are the same as in the formulas for self-inductance already given.

For two parallel coaxial multilayer coils of square or nearly square cross section, a good approximation is given by

$$M = n_1 n_2 M_0 \quad (27)$$

where n_1 and n_2 are the numbers of turns on the two coils, and M_0 is the mutual inductance of two circles located at the centers of the cross sections of the two coils.

The same formula may be used as a rough approximation for the mutual inductance of two coaxial single-layer solenoids.

References

- Butterworth, S.: The High-frequency Copper Losses in Inductance Coils, *Exptl. Wireless*, **2**, 613, 1925.
 Coursey, P. R.: Calculation and Design of Inductances, *Electrician*, **75**, 841, 1915.
 High-frequency Resistance and Capacitance of Single-layer Solenoids, *Wireless Engr.*, February and March, 1940.
 Griffiths, W. H. F.: Notes on Standard Inductances for Wave Meters and Other Radio-frequency Purposes, *Exptl. Wireless and Wireless Engr.*, **6**, 543, 1929.
 Kantor, Myron: Theory and Design of Progressive and Ordinary Windings, *Proc. IRE*, **35**, 1563, December, 1947.
 Pernot, F. E.: Formulae and Tables for Design of Air-Core Inductance Coils, *Univ. Calif. Pub., Eng.*, **1**, 117, 1916.
 Soucy, C. I.: Temperature Coefficient Effects of R-f Coil Finishes, *Tele. Tech.*, **6**, 53, December, 1947.
 Whittemore, L. E., and G. Breit: Inductance, Capacity, and Resistance of Coils at Radio Frequency, *Phys. Rev.*, **14**, 170, 1919.
 Wilmotte, R. M.: Parasitic Losses in Inductance Coils at Radio Frequency, *Exptl. Wireless*, **2**, 451 and 477, 1925.

31. Coil Form Relations. From the preceding data it will be seen that for solenoid windings the maximum inductance occurs when the diameter of the winding divided by the length of the winding equals 2.46. For maximum inductance in multilayer solenoids, the winding cross section should be square and the diameter of the form should be about equal to twice the length of a side of that square. For solenoids of the same inductance, the distributed capacitance is approximately proportional to the radius of the form and independent of the number of turns and length of winding.

32. Choice of Wire Size. The choice of wire size for universal windings or progressive universals depends on the desired Q , the permissible outside diameter, and the cost. At the frequencies at which these windings are used, roughly 200 to 2,000 kc stranded wire is desirable to give good Q . In general for a given inductance, the greater the number of strands, the higher the cost and the better the Q . The smaller the strand size, the higher the Q and the cost. The smaller the strand size, the less the outside diameter, a desirable relation with iron cores or where shielding is to be used. The strands should be loosely twisted, about 10 or 12 twists per foot for best results. The effect of wire size on Q , relative cost, and relative outside diameter for a 1-mh winding using some common wire sizes is given in Table 7.

In the region from about 2 to 6 Mc, solenoid windings are used with a wire size of from No. 30 to No. 34. From 6 to 100 Mc, the common wire sizes are from No. 18 to No. 24.

33. Choice of Wire Insulation. There are many insulating materials now in use on wire. These fall into two basic groups; the filament type and the extruded or dipped type. Silk, nylon, cotton, rayon, and glass are filament types. Of the latter type are enamel, polyvinyl formal (Formex or Formvar), nylon, and polyethylene. Since the insulation becomes a large part of the dielectric for the distributed capacitance, it is important in high- Q coils to choose insulators with low dielectric losses. Furthermore, in many cases the thickness of the insulation controls the spacing between wires and is, therefore, important. The filament or thread type of insulators is frequently put on over an enameled base to prevent short circuits between turns in case the

threads should separate. Polyvinyl formal and nylon coatings have extremely high abrasion resistance and are useful where the winding may be subjected to severe handling or where the movement of the wire with respect to an abrasive surface, such as an iron core, might cause trouble. The only one of the dipped or extruded coatings that has really low losses is polyethylene. This material has not yet seen wide enough usage to predict its full possibilities accurately.

Cotton gives a very thick insulation and is not too commonly used owing to the large amount of space occupied. It is frequently used for broadcast loop antennas. By far the most common filament insulations are silk and nylon. Nylon is not quite so satisfactory as silk owing to its poor power factor, especially when the temperature is raised slightly.

It should be remembered that the insulation on wire is a necessary evil. It should be omitted wherever possible, *e.g.*, in spaced solenoids wound on threaded forms. In any case it should be kept to the practical minimum.

34. Choice of Form. Form size is dictated by space limitations, winding type, winding design, and cost. The important factor left to consider is material. It must be strong enough to support the winding and, since it is usually largely within the field of the winding, it must be electrically good. It should be dimensionally stable with temperature changes. No material yet known is electrically perfect and, therefore, the less form material used, the less additional losses will be introduced.

Common materials are ceramics, such as steatite, thermosetting plastics, such as molded or laminated phenolics, thermoplastics, such as polystyrene and its copolymers, cellulose acetate and its copolymers, ethyl cellulose, glass fiber, and paper.

Ceramics and glass are good electrically but are fragile and cannot be made to close dimensional tolerances. As a group they are the most expensive. The phenolics vary widely in their electrical properties depending on the filler used. The better electrical grades are satisfactory and are mechanically better and less costly than ceramics. Certain of the plastics, particularly the styrenes and ethyl cellulose, have the best electrical properties but are not too widely used because they will not stand the rather high temperatures encountered in modern receivers. This is particularly true when the winding has appreciable tension.

Fiber and paper are economical. When properly treated with wax, these materials are suitable for all but the most exacting applications.

Certain windings are completely self-supporting and require no form. Others are wound directly on powdered-iron cores.

35. Choice of Finish. The finish on an inductor performs two major functions: it helps to make the unit mechanically stable and to make it electrically stable through protection against ambient conditions. The most common finishes are waxes, varnishes, and lacquers, or synthetics. The choice of finish depends on the conditions to be encountered and on the amount of additional losses that can be tolerated in the final design. The use of any finish always introduces some loss.

Operating temperature ranges may be from -20 to $+160^{\circ}\text{F}$ for civilian broadcast units to -70 to $+185^{\circ}\text{F}$ for air-borne military equipment. Relative humidities may also vary over a wide range.

Waxes, particularly the microcrystalline type, ceresins, etc., are excellent for the moderate temperatures of civilian use. They are simple to use, are inexpensive, have fairly low losses, and low dielectric constants, are noncorrosive, and provide the best humidity protection. They should never be used when the ambient temperature is close to their melting point.

Varnishes, both the solvent and solventless types, are usually not so good electrically, particularly at the higher frequencies. The initial expense and cost of application are high. The common heat polymerizing varnishes have a short shelf life before use. However, properly applied varnishes give good humidity protection, withstand a greater temperature range than waxes, and are mechanically strong.

Lacquers or synthetics are easier to apply than varnishes, give about as good mechanical strength, have lower losses, but do not give so good humidity protection.

They can be made so that they have low dielectric constants. This fact is important since any finish becomes a part of the dielectric of the distributed capacitance; in many designs it is a large part.

In choosing a finish the following points are important: (1) the temperature range. The finish must hold the windings firmly throughout the operating temperature range. It must not change physically with temperature changes for that would distort the winding. It must not change its dielectric constant since that would change the apparent inductance. (2) The finish should be as good a dielectric as possible so that the losses will be kept to a minimum. The dielectric constant must be low to keep the distributed capacitance low. (3) The finish should afford good humidity protection and at the same time be noncorrosive. Since many inductors have d-c potentials applied and are wound with fine wire or stranded wire, any tendency for corrosion would be accelerated and quickly result in an open circuit.

Accelerated corrosion tests can be made with typical universal windings spaced $\frac{1}{4}$ in. apart, wound with 7/41 SSE litz wire, subjected to 90 to 95 per cent relative humidity at 100 to 120°F, with 400 volts d.c. between the windings. Under these conditions, windings treated with practically any lacquer or varnish show at least one and at most seven strands of the litz completely corroded after 96 hr. Wax-treated units in the same test rarely show more than one broken strand and usually not any.

36. Methods of Obtaining Close Coupling. For universal windings, close coupling is obtained by making the second winding a continuation of the first, sometimes separated by a layer of tape. The two windings may be placed side by side very close together. In some cases maximum coupling has been obtained by bifilar winding in which two conductors are wound simultaneously and are always parallel as the winding is built up.

For solenoids bifilar winding is also used. Less coupling is obtained by interwinding only a part of the two windings. The second winding may be placed directly over the first, separated by a tape.

Similar methods apply to other types of windings. Where powdered-iron cores are used, close coupling can be obtained by simply winding both coils on the same iron. Such practice is not recommended unless high-resistance iron is used.

Coupling as high as 98 per cent can be obtained with bifilar universals and as high as 90 per cent with bifilar solenoids. The use of iron cores in the windings and iron cups surrounding the windings will increase the coupling in most cases.

37. Effect of Shielding. To prevent undesirable coupling between inductors it is common to surround them with a shield of conducting material. The best conductor is copper and, although it is sometimes used, it is expensive and requires some treatment to maintain its appearance. Easy to fabricate, light in weight, low in cost, and requiring no treatment, aluminum is a favorite shield material. Zinc is also used, especially when it is necessary to soft-solder to the shield.

Shielding is accomplished because of the eddy currents set up in the conducting material. The eddy currents oppose the inducing flux, and therefore, repel and confine it. This opposition to the flux acts as a negative permeability of the surrounding medium. Because the shield is not a perfect conductor, losses are introduced that reflect as losses in the inductor. Shielding lowers the inductance and the Q . The proximity of the shield to the inductor and the shield material controls the amount of lowering that takes place.

Shielding is also accomplished through the use of cups or sleeves of powdered iron. In this case the flux is concentrated in the low-reluctance path of the iron. The Q and inductance of the winding may be slightly increased. This type of shielding is not complete, and usually the whole is surrounded with a conventional aluminum shield, which then has little effect on the winding. Its function is to provide electrostatic shielding for more complete shielding.

38. Special Considerations in the Use of Powdered-iron Cores. The use of powdered-iron cores materially changes some details of winding design. It is most

important to get as much of the winding as possible as near the core as possible. In so doing the maximum number of turns of the conductor is affected by the core. In other words, there is less leakage flux and more active flux to produce the maximum inductance.

In the case of universal windings, the depth of the winding should be kept to a minimum to keep the average turn closer to the core. In all cases, the less the wall thickness of the form used, the better the results.

39. Typical Inductors Currently in Use. The following inductors are used in a typical broadcast receiver employing a loop antenna. The set has a standard 365- μf three-gang variable capacitor. All inductors are impregnated and flashed in wax. Shields used for r-f coil, first and second i-f coils, are all $1\frac{1}{4}$ in. square aluminum, $2\frac{1}{2}$ in. long.

Loop Antenna.

Wound in the form of a flat spiral, elliptical in shape, minor diameter of average turn = 5 in., major diameter of average turn = 11 in., wire = 22 DCE, Q at 1 Mc = 149 far from chassis and 72 mounted on chassis, turns = 27, inductance = 265.7 μh , $C_d = 16.8 \mu\text{f}$.

R-f Coil.

Type	Turns	Width, in.	Wire	Air inductance	Q in shield at 1 Mc
Primary universal.....	646	$1\frac{3}{64}$	39 SSE	6.38 mh	98
Secondary 4-layer hank.....	195	$\frac{3}{41}$ SSE	280 μh	

Spacing primary to secondary $\frac{5}{64}$ in. $L_M = 200 \mu\text{h}$. Form: $\frac{1}{2}$ in. OD by $\frac{3}{16}$ in. ID, bakelite.

Oscillator Coil.

Type	Turns	Width, in.	Wire	Air inductance, μh
Primary universal.....	23 $\frac{3}{4}$	$\frac{3}{16}$	36 SCE	14
Secondary universal.....	83 $\frac{1}{4}$	$\frac{3}{16}$	36 SCE	127

Secondary wound on top of primary separated by yellow cambric $\frac{3}{8}$ by 0.010 by $2\frac{1}{16}$ in. $L_M = 25.1 \mu\text{h}$ in air. Form: $\frac{1}{2}$ in. OD by $\frac{3}{8}$ in. ID, wax-impregnated paper.

First I-f Transformer.

Type	Turns	Width, in.	Wire	Air inductance, mh	Q in shield at 455 kc
Primary universal.....	422	$\frac{3}{16}$	$\frac{5}{42}$ SSE	2.27	83
Secondary universal.....	422	$\frac{3}{16}$	$\frac{5}{42}$ SSE	2.27	82

Spacing primary to secondary $\frac{9}{16}$ in. $L_M = 109.2 \mu\text{h}$ in air. Form: $\frac{3}{8}$ in. OD by $\frac{5}{16}$ in. ID, wax-impregnated paper.

Second I-f (Diode) Transformer.

Type	Turns	Width, in.	Wire	Air inductance, mh	Q in shield at 455 kc
Primary universal.....	365	$\frac{3}{16}$	$\frac{5}{42}$ SSE	1.7	90
Secondary universal.....	415	$\frac{3}{16}$	$\frac{5}{42}$ SSE	2.2	83

Spacing, primary to secondary, $3\frac{3}{4}$ in. $L_M = 130.7 \mu\text{h}$ in air. Form: $\frac{3}{8}$ in. OD by $\frac{5}{16}$ in. ID, wax-impregnated paper.

The following inductors are used in a typical permeability-tuned automobile broadcast receiver covering 535 to 1,610 kc.

Antenna and R-f Coils.

Form: 0.223 in. OD by 0.207 in. ID, bakelite. 356 turns 5/43 SSE straight progressive universal winding $1\frac{1}{4}$ in. long. Minimum $Q = 60$ at 600 kc. Core: 0.200 in. OD by $1\frac{1}{2}$ in. long. Apparent permeability = 9. Lacquer-coated. These coils tune to 1,610 kc with $65 \mu\text{mf}$, no core.

Oscillator Coil.

Form: 0.318 in. OD by 0.207 in. ID, bakelite. 140 turns 36 E. Variable-pitch solenoid, $1\frac{1}{2}$ in. long. Core same as for antenna and r-f coils. Lacquer-coated. This coil tunes to 1,870 kc with $196 \mu\text{mf}$, no core. Cathode coil wound on $\frac{7}{16}$ in. OD by $\frac{3}{8}$ in. ID bakelized paper. Winding: 16 turns of No. 30 E at 13 turns per inch. This coil is slipped over the oscillator winding and exactly centered over it.

First I-f 260-kc Transformer.

Type	Width, in.	Wire	Turns	Air inductance, mh	Q in shield at 250 kc
Primary universal.....	$\frac{3}{4}$	38 SSE	390	2.0	59
Secondary universal.....	$\frac{3}{4}$	38 SSE	390	2.0	59

Form: $\frac{3}{8}$ in. OD by $\frac{5}{16}$ in. ID, wax-impregnated paper. Spacing primary to secondary $1\frac{1}{2}$ in. $L_M = 89 \mu\text{h}$ in air.

Second I-f (Diode) 260-kc Transformer.

Type	Width, in.	Wire	Turns	Air inductance, mh	Q in shield at 260 kc
Primary universal.....	$\frac{3}{4}$	38 SSE	542	4.0	55
Secondary universal.....	$\frac{3}{4}$	38 SSE	340	1.5	51

Form: $\frac{3}{8}$ in. OD by $\frac{5}{16}$ in. ID, wax-impregnated paper. Spacing primary to secondary $2\frac{3}{4}$ in. $L_M = 156 \mu\text{h}$ in air.

MAGNETIC MATERIALS¹

BY RAYMOND L. SANFORD²

Magnetic materials are essential in the construction of nearly all types of electrical machinery and apparatus. In the field of radio communication, they are employed in the cores of transformers, choke coils, relays, etc., and for magnetic shielding. Permanent magnets are used in d-c measuring instruments, polarized relays, magnetrons, and loud-speakers. A great variety of magnetic materials is produced commercially, and new types are continually being developed. In view of the diversity in requirements for practical applications, a knowledge of the magnetic characteristics of available materials is needed in connection with the design of various types of apparatus. The principal producers of magnetic materials issue technical data giving the properties of the materials they make. Before undertaking the design of apparatus involving the use of magnetic materials it is best to obtain up-to-date information from one or more of these manufacturers.

¹ The material for this portion of Chap. 3 came in part from "Permanent Magnets" and "Magnetic Testing," Nat. Bur. Standards Circ., C448 and C456.

² National Bureau of Standards, Washington, D.C.

MAGNETIC QUANTITIES AND UNITS

The characteristics of magnetic circuits and the magnetic materials of which they are made are expressed in terms of certain magnetic quantities and units which may be described as follows:

40. Magnetomotive Force. Magnetization is generally accomplished by means of electric current in windings linked with a magnetic circuit of which the material to be magnetized forms a part. The total measure of the magnetizing effect of such a coil is called the magnetomotive force, \mathfrak{F} . The cgs unit of magnetomotive force (mmf) is called the *gilbert* and is defined by the equation

$$\mathfrak{F} = 0.4\pi NI$$

where \mathfrak{F} = mmf, gilberts

N = number of turns in the coil

I = current, amp

41. Magnetic Flux. The total measure of the magnetized condition of a magnetic circuit when acted upon by a magnetomotive force is called the "magnetic flux," ϕ . It is characterized by the fact that a variation in its magnitude gives rise to an emf in an electric circuit linked with it. The emf thus induced is at any instant directly proportional to the time rate of variation of the flux. The cgs unit of magnetic flux is called the *maxwell* and is defined by the equation

$$e = N \left(\frac{d\phi}{dt} \right) 10^{-8}$$

where e = induced emf, volts

N = number of turns linked with the flux

ϕ = magnetic flux, maxwells

t = time, sec

42. Magnetic Reluctance. That property of a magnetic circuit which determines the relationship between the magnetic flux and the corresponding mmf is called the "magnetic reluctance," \mathcal{R} , of the circuit. The cgs unit* is defined by the equation

$$\phi = \frac{\mathfrak{F}}{\mathcal{R}}$$

where ϕ = magnetic flux, maxwells

\mathfrak{F} = mmf, gilberts

\mathcal{R} = magnetic reluctance in cgs units

In a magnetic circuit of uniform cross-sectional area and uniform permeability (to be defined later),

$$\mathcal{R} = \frac{l}{\mu A}$$

where \mathcal{R} = magnetic reluctance in cgs units

μ = magnetic permeability

l = length, cm

A = cross-sectional area, sq cm

43. Magnetic Permeance. Magnetic permeance, \mathcal{P} , is the reciprocal of the magnetic reluctance. Thus,

$$\phi = \mathfrak{F} \mathcal{P}$$

where ϕ = magnetic flux, maxwells

\mathfrak{F} = mmf, gilberts

\mathcal{P} = magnetic permeance in cgs units

In a circuit of uniform cross section and permeability

$$\mathcal{P} = \frac{\mu A}{l}$$

* The unit of magnetic reluctance was called the "oersted" until 1930, when the International Electrotechnical Commission adopted that name for the unit of magnetizing force, thus leaving the unit of reluctance without a name.

that their effect is added to that of the applied field. The flux density due to the combined effect of the applied field and the domains is the magnetic induction, B , and that due to the domains alone is the intrinsic induction, B_s .

Upon removal of the magnetizing force, the orientations of the domains resulting from the applied field tend to persist to a greater or less extent, depending upon their distribution in space, which in turn depends upon the internal structure of the material. To reduce the induction further, it is necessary to apply a reverse, or demagnetizing, force. The relationships between induction and the applied field are shown graphically in Fig. 12, where the abscissas represent values of magnetizing force, H , and the ordinates are values of magnetic induction, B . The curve $oabc$ is called the "normal induction curve" and shows how the induction varies as the magnetizing force applied to a material initially in the demagnetized condition is increased from zero. At any point on this curve, the ratio of the value of induction to the corresponding value of magnetizing force is called the *magnetic permeability*, μ . Thus the coordinates of point a on the curve are B_a and H_a , respectively, and the permeability is

$$\mu = \frac{B_a}{H_a}$$

Since the line $oabc$ is not straight, the permeability is not constant but varies with the degree of magnetization. This variation is one of the distinguishing characteristics of ferromagnetic materials.

48. Hysteresis Effects. If, after the magnetizing force has reached a certain value (as at point b), it is then decreased, the induction does not follow the curve $oabc$ in reverse order but lags behind the magnetizing force, as shown by the curve bd . Thus, when the magnetizing force has been reduced to zero, point d , the induction still has an appreciable value. This value is called the *residual induction*, B_r .

To reduce the induction still further, it is necessary to increase the magnetizing force in the opposite direction. The value of the reversed magnetizing force required to reduce the induction to zero is called the *coercive force*, H_c .

The lagging of the induction behind the magnetizing force is called *hysteresis*, and the complete curve $bdefghb$, is called a hysteresis "loop." Hysteresis is another of the characteristics of ferromagnetic materials and is the property that makes permanent magnets possible. The size of the hysteresis loop, and consequently the magnitudes of B_r and H_c , depends upon the values of B and H at the tip of the loop, generally designated by the symbol B_m and H_m . As H_m is increased, the size of the hysteresis loop approaches a maximum. The values of B_r and H_c corresponding to the maximum loop for a given material are called the *retentivity* and *coercivity*, respectively.

A major hysteresis loop is produced by varying the magnetizing force continuously from a positive maximum value to the same value negative and back again to the positive maximum. However, if a change in direction of the variation is made at some intermediate point on the major loop, a minor loop is produced, as shown between the points i and k in Fig. 12. The slope of the line joining the tips of such minor loops is called the *incremental permeability*, μ_Δ . In other words, the incremental permeability is the ratio of the change in B to the corresponding change in H when the average value of B differs from zero. That is, $\mu_\Delta = \Delta B / \Delta H$.

That part of the hysteresis loop extending from the residual induction, B_r , to the coercive force, H_c , *i.e.*, from d to e , is called the *demagnetization curve*. Points on this curve are designated by the coordinates B_d and H_d . Most of the important characteristics of a permanent-magnet material can be indicated by points on this curve or within the area between it and the axes of coordinates. The value of H_m necessary to produce the maximum loop depends upon the type of material and ranges from about 300 oersteds for most of the older types of material to about 3,000 oersteds or more for the newer types. Figure 13 represents a typical demagnetization curve, together with the so-called *energy-product* curve obtained by plotting the product of corresponding values of B_d and H_d on the demagnetization curve against B_d . The

maximum value of this product, $(B_d H_d)_{max}$, can be considered as a criterion of quality and is useful for the comparison of various materials. It is generally not necessary to plot the $B_d H_d$ curve to determine the maximum energy product. A line drawn from the origin of coordinates to the intersection of horizontal and vertical lines through B_r and H_c , respectively, intersects the demagnetization curve at the point where the product $B_d H_d$ is a maximum for most materials.

MAGNETIC CHARACTERISTICS OF MATERIALS

Magnetic materials employed in commercial practice may be considered under the following classifications: (1) solid-core materials, (2) electrical sheet and strip, (3) special-purpose alloys, (4) permanent-magnet alloys.

49. Solid-core Materials. These materials are used for the cores of d-c electromagnets, relays, field frames of d-c machines, etc. The principal requirement is high permeability particularly at relatively high induction. For most uses it is also desirable that the hysteresis be low. The principal materials employed are soft iron, relay steel, cast steel, cast iron, and an alloy of approximately 35 per cent of cobalt and 65 per cent of iron known as ferrocobalt. Ferrocobalt is characterized by very high permeability in the upper part of the normal induction range and a saturation induction approximately

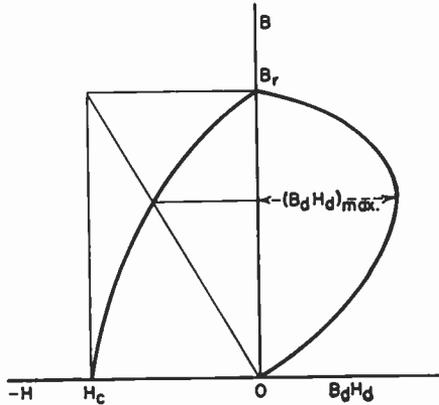


FIG. 13. Demagnetization and energy-product curves.

10 per cent greater than that of pure iron. Its cost is relatively high, however, and its use is limited in general to pole pieces in which a very high induction is required. Cast iron has a relatively low permeability and is used principally in field frames when cost is of primary importance and extra weight is not objectionable. Several varieties of soft iron are available, such as Norway iron, Armco iron, and Swedish charcoal iron. Relay steels contain from 1 to 3.25 per cent of silicon to reduce aging. Electrolytic iron may also be used. All these materials are relatively pure iron, low in carbon and other impurities. Cast steel should be low in carbon, not over 0.1 to 0.2 per cent, and contain only the usual small amounts of the ordinary impurities. Cast iron is high in carbon, about 3 per cent, and also contains about 3 per cent of silicon, and varying percentages of phosphorus, manganese, and sulfur.

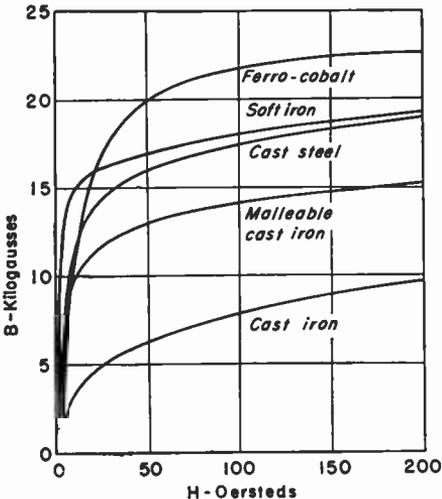


FIG. 14. Typical normal induction curves for solid-core materials.

The best magnetic properties are obtained by a suitable annealing treatment. The properties of cast iron can be greatly improved by malleablizing, a process that converts a large part of the carbon to the amorphous form.

50. Electrical Sheet and Strip. The terms *electrical sheet* and *electrical strip* are commonly used to designate silicon-iron alloys produced in sheet or strip form and used as core materials in a-c apparatus, such as transformers, generators, motors, electromagnets, or relays. The principal requirements are high permeability, low hysteresis, and high resistivity. The several grades differ mainly with respect to their silicon content, which ranges from about 0.5 per cent to approximately 4.5 per cent. Alloys containing the higher percentages of silicon are practically nonaging; *i.e.*, the permeability and losses do not change with time. The required magnetic properties are produced by annealing.

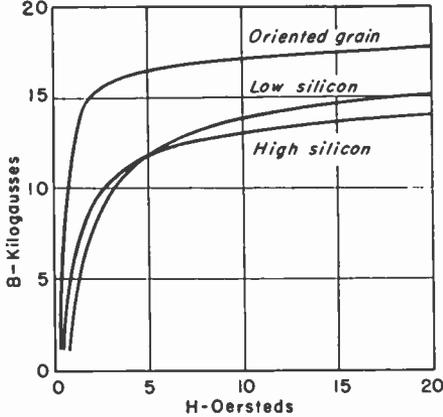


FIG. 15. Typical normal induction curves for electrical sheet.

By a suitable combination of cold-rolling and heat-treatment, materials are produced in which the crystal axes are given a definite orientation. Such material has considerably better properties, when magnetized in the preferred direction, than the ordinary grades. Figure 15 shows typical normal induction curves for two grades of electrical sheet and orientated-grain material. The improvement in the orientated-grain material is particularly conspicuous in the upper part of the normal induction curve.

The different grades of electrical sheet and strip are usually sold on the basis of guaranteed maximum values of total core loss, as determined in accordance with the specifications of the ASTM.²

The common designations of the various grades are armature, electrical, motor, dynamo, and transformer. The transformer grades are further subdivided into classes denoted by numerals corresponding to the core loss under standard conditions. Armature, electrical, and motor grades are used principally in small motors, a-c magnets, and starting transformers. The dynamo grade is used in high-efficiency rotating machines and small transformers. The transformer grades are used in power and radio transformers (see Table 12).

Table 12. Typical Characteristics of Commercial Electrical Sheet*

Grade	Si per cent	Core loss†		Max permeability		Coer-cive force, H _c	Resistivity, microhm-cm	Density, g/cm ³
		B _m = 10,000	B _m = 15,000	μ _m	B at μ _m			
Armature.....	0.5	1.30	3.06	5,800	7,000	0.70	19	7.83
Electrical.....	1.0	1.17	2.75	6,150	6,000	0.68	26	7.79
Motor.....	2.5	1.01	2.37	6,100	6,100	0.58	42	7.70
Dynamo.....	3.25	0.82	1.93	5,800	5,000	0.50	50	7.65
Transformer 1.....	4.0	0.72	1.69	6,300	5,000	0.40	56	7.60
Transformer 2.....	4.5	0.58	1.36	8,300	4,500	0.25	62	7.57
Oriented grain‡.....	3.3	0.04	0.09	40,000	6,000	0.10	48	7.65

* Subject to some variation in actual practice.

† Watts per pound at 60 cycles.

‡ Flux in direction of rolling.

51. Special-purpose Alloys. For certain applications, special alloys have been developed which, after proper heat-treatment, have superior properties in certain ranges of magnetization. For instance, alloys of nickel and iron with possible small

percentages of molybdenum or chromium have very high values of initial and maximum permeability. Alloys of this class, which may have from 70 to 80 per cent of nickel, are called "permalloys." An alloy of 50 per cent of nickel and 50 per cent of iron is called "hipernik." Another alloy having a small percentage of copper in its composition is called "mumetal." The characteristics of these alloys differ in detail, but in general they have high initial and maximum permeability, low hysteresis, and low saturation values. They are particularly applicable for use at low inductions. Typical permeability curves are given in Fig. 16.

A certain alloy of nickel, cobalt, and iron after suitable heat-treatment has very nearly constant permeability for inductions below 1,000 gauss and is called *perminvar*. The 50-50 nickel-iron alloy can also be heat-treated so as to have similar characteristics.

An alloy of equal proportions of iron and cobalt has high permeability which persists at higher values of induction than the nickel-iron alloys and is called *permendur*.

Typical characteristics of some special magnetic alloys are given in Table 13.

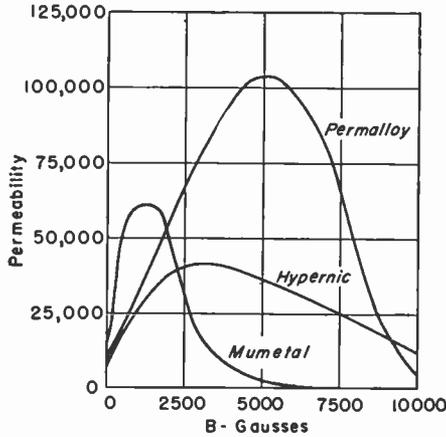


FIG. 16. Typical permeability curves for special-purpose alloys.

Table 13. Typical Characteristics of Some Special Magnetic Alloys*

Alloy	Nominal composition, per cent †	Initial permeability, μ_m	Max permeability, μ_o	Coercive force, H_c	Saturation induction, B_s	Resistivity, microhm-cm	Density, g/cm ³
Superalloy.....	79Ni, 5Mo	100,000	800,000	0.004	8,000	60	8.87
78 permalloy.....	78Ni	8,000	100,000	0.05	10,000	16	8.60
4-79 permalloy.....	79Ni, 4Mo	20,000	80,000	0.05	8,700	57	8.72
45 permalloy.....	45Ni	2,500	25,000	0.3	16,000	50	8.17
Hipernik.....	50Ni	4,000	80,000	0.05	16,000	35	8.25
4750 alloy.....	47 to 50Ni	5,000	32,000	0.03	16,000	45	8.15
1040 alloy.....	72Ni, 14Cu, 3Mo	40,000	100,000	0.02	6,000	56	8.76
Mumetal.....	75Ni, 2Cr, 5Cu	20,000	110,000	0.02	7,200	42	8.60
Permendur.....	50Co	800	5,000	2.0	24,500	7	8.30
Powdered permalloy...	81Ni, 2Mo	125	130	10 ⁴	7.8

* Subject to some variation in actual practice.

† Balance iron plus usual small amounts of impurities.

Another series of magnetic alloys of copper, nickel, and iron is temperature-sensitive, having an approximately linear relation between permeability and temperature. These are called "thermalloys." The principal use is in the compensation of watt-hour meters for temperature variations. They are also used in certain types of thermal relays.

52. Permanent-magnet Alloys. For permanent magnets, high values of residual induction and coercive force are required. Ordinary high-carbon steels and some special tool steels make fair magnets when properly hardened, but better results are obtained by the use of alloys specially made for the purpose. Development along this line has been rapid during the past few years and is still going on. There are at present three general classes of permanent-magnet materials: those in which the principal

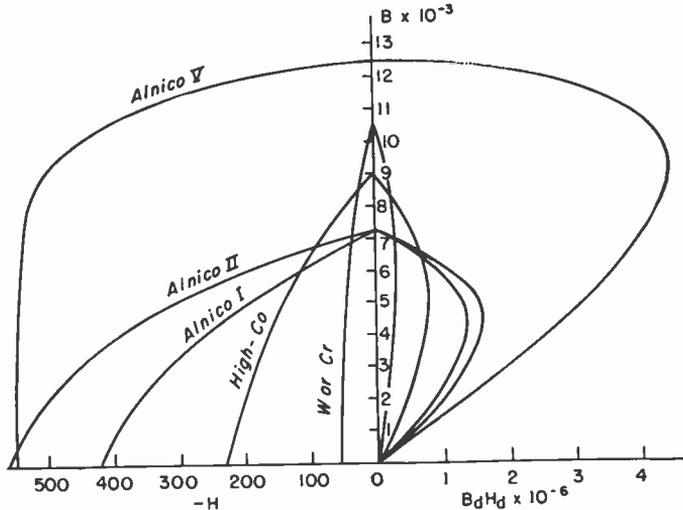


FIG. 17. Demagnetization and energy-product curves for typical magnet materials.

alloying elements are tungsten or chromium or both; those in which cobalt is present in substantial amounts; and those in which aluminum and nickel play an important part.

Tungsten, chromium, or cobalt magnet steels are hardened by quenching. The aluminum-nickel alloys are among the so-called dispersion-hardening alloys, the final

Table 14. Typical Characteristics of Several Permanent-magnet Alloys^a

Type	Nominal composition ^b	Residual induction, B_r	Coercive force, H_c	(Ball) ^c $\times 10^{-4}$	Fabrication ^c
Tungsten.....	5.5W, 0.6C	10,000	65	0.27	Hot-forge or cast, machine
3½% Cr.....	3.5Cr, 0.3Mn, 0.9C	9,500	63	0.25	Hot-forge or cast, machine
Low Co.....	8.5Co, 5Cr, 1W, 0.9C	7,500	120	0.38	Hot-forge or cast, machine
Co-Cr.....	16Co, 9Cr, 1C	8,000	180	0.61	Hot-forge or cast, machine
High Co.....	36Co, 5W, 2Cr, 0.8C	9,000	230	0.87	Hot-forge or cast, machine
Alnico 1.....	12Al, 20Ni, 5Co	7,300	440	1.40	Cast, grind
Alnico 2.....	10Al, 17Ni, 12.5Co, 6Cu	7,350	560	1.60	Cast or sinter, grind
Alnico 3.....	12Al, 25Ni	6,900	475	1.38	Cast, grind
Alnico 4.....	12Al, 28Ni, 5Co	5,300	730	1.30	Cast, grind
Alnico 5.....	8Al, 14Ni, 24Co, 3Cu	12,500	640	5.00	Cast, grind ^d
Alnico 12.....	6Al, 18Ni, 35Co, 8Ti	5,700	900	1.80	Cast, grind
Comol.....	12Co, 17Mo	10,500	250	1.10	Cast or hot-roll and punch
Cunife 1.....	60Cu, 20Ni	5,800	600	1.96	Cold-roll, machine ^e
Cunife 2.....	50Cu, 20Ni, 2.5Co	7,300	260	0.78	Cold-roll, machine ^e
Cunico.....	50Cu, 21Ni, 29Co	3,400	710	0.85	Cast, cold-roll, machine
Vicalloy.....	52Co, 9.5V	9,600	500	2.80	Cast, hot-swage, cold-draw
Silmaual.....	87Ag, 9Mn, 4Al	500	5,500 ^f	0.06	Machine-punch, cold-roll
Vecolite.....	30Fe ₂ O ₃ , 40Fe ₃ O ₄ , 26CO ₂ O ₃	1,600	900	0.50	Sinter ^g

^a To be taken as typical only, compositions and properties vary somewhat in practice.

^b Balance iron plus usual small amounts of impurities.

^c Hardened at proper stage either by manufacturer or in accordance with manufacturer's instructions

^d Heat-treated in a magnetic field; has directional properties.

^e Has directional properties; should be magnetized in direction of rolling

^f Intrinsic coercive force, H_i .

treatment for which is prolonged heating at a temperature of the order of 660 to 700°C. The alloys that are hardened by quenching can be forged and machined, but most of the dispersion-hardening alloys cannot be forged or machined and must be cast to shape and can be cut only by grinding.

The important magnetic characteristics of permanent-magnet alloys are indicated by the part of the hysteresis loop lying between the residual induction and the coercive force and a curve plotted between B and the products of corresponding values of B and H for the same part of the loop. These products are proportional to the magnetic energy stored at the given values of induction. The maximum value of the energy product may be taken as an index of magnetic quality. Typical curves for the three classes of permanent-magnet material are shown in Fig. 17.

The choice of material for a given application depends upon several factors and, consequently, no one material can be considered as best under all conditions. (See Table 14 for typical characteristics of several permanent-magnet alloys.)

MAGNETIC TESTING

53. Tests with D.C. Most commercial magnetic testing in the United States is done in accordance with the specifications of the ASTM.³ The methods or apparatus currently approved by the ASTM include the Rowland ring method,⁴ the Burrows compensated double yoke,⁵ the Fahy simplex permeameter,⁶ the high-H permeameter,⁷

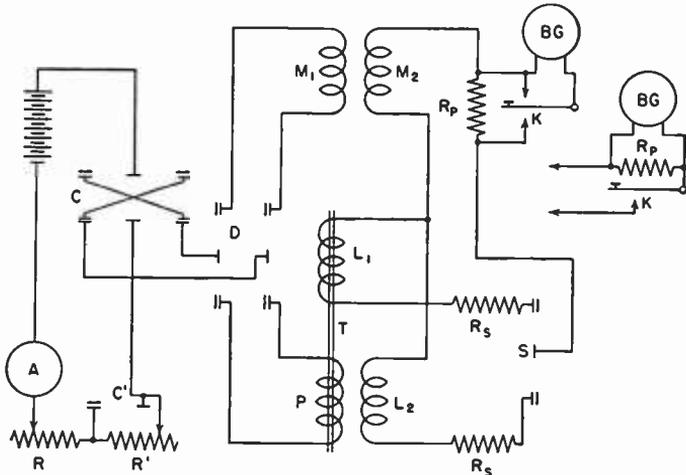


FIG. 18. Connections for ballistic test.

the Babbitt permeameter,⁸ the simplex super-H adapter,⁹ and the saturation permeameter.¹⁰ Many makers of magnetic materials employ, for purposes of quality control, special methods or apparatus that are more simple and rapid and give comparative values sufficiently accurate for the purpose.

Ballistic methods, so-called because they employ a ballistic galvanometer in the measurements, are most commonly used in ordinary magnetic testing.

The principles involved in all ballistic methods are the same, the differences being in the type of magnetic circuit and the arrangement of the magnetizing and test coils. A typical diagram of connections is shown in Fig. 18. Current from the storage battery is controlled by R and R' and measured by ammeter A , or, if desired, by a standard shunt and a potentiometer. C is a reversing switch, and switch C' serves to insert the auxiliary resistance R' into the circuit. The current flows in primary M of a mutual inductor when switch D is closed upward and in the magnetizing windings of the apparatus P , when D is closed downward.

The secondary M_2 of the mutual inductor, or an equivalent noninductive resistance, is always included in the galvanometer circuit. By means of selector switch S , the galvanometer can be connected either to test coil L_1 , which surrounds test specimen T , or to coil L_2 , which does not surround the specimen but is located in such a position as to give an indication of the magnetizing force. The sensitivity and damping of the galvanometer are adjusted by means of the parallel resistance R_p and the series resistances R_s . In some types of apparatus, magnetizing force is measured in terms of the magnetizing current, in which case coil L_2 and the resistance in series with it are omitted. Two alternative methods of galvanometer control are shown. The choice between these two methods is mainly a matter of individual preference.

As shown in the main diagram, the galvanometer can be short-circuited, open-circuited, or connected to the test circuit by closing the key upward, opening it, or closing it downward, respectively. By the alternative connection shown at the right the parallel resistance, usually of such a value as to give critical damping, is always connected to the galvanometer, which in turn is connected to the test circuit or not according as the key is closed or open. If desired, a second parallel resistance may be connected between the lower contact of the key and the left-hand terminal of the galvanometer. With this connection, the galvanometer will be overdamped when connected to the test circuit.

The galvanometer is calibrated by means of a standard mutual inductor. It is usually convenient to adjust the sensitivity so that the scale is direct reading in terms either of induction or of magnetizing force, thus avoiding the necessity of multiplying the scale reading by odd factors. It is customary to make a deflection of 1 cm correspond to the reversal of an induction of 1,000 gausses or of a magnetizing force of 1, 10, or 100 oersteds as required. When calibrating for induction, the current to be reversed in the primary of the mutual inductor depends upon the value of the mutual inductance, the number of turns in the test coil, and the cross-sectional area of the specimen, and is calculated from the formula:

$$I_c = \frac{BAN}{M} \times 10^8$$

where I_c = calibrating current, amp

B = induction, gausses

A = cross-sectional area of specimen, sq cm

N = number of turns in the test coil

M = mutual inductance, henrys

The corresponding formula for calibrating for magnetizing force is

$$I_c = \frac{HAN}{M} \times 10^8$$

where H = magnetizing force, oersteds

AN = product of the number of turns by the average area of the test coil

The other quantities are the same as in the preceding formula.

Since the calibration is made by reversal of the calibrating current, care must be taken when measuring changes in induction or magnetizing force as in the determination of points on the hysteresis loop to multiply the readings by 2.

Unless the test coil for the measurement of induction is wound directly on the specimen, a correction must be made to the observed value of induction. This is to take into account the flux in the space between the specimen and the test coil. The amount to be subtracted from the observed value of induction is proportional to the magnetizing force and depends upon the difference between the areas of the test coil and the specimen. The correction is equal to kH , where $k = (a - A)/A$, a = area of test coil, and A = area of specimen.

In making a normal-induction test, the specimen is first demagnetized. The magnetizing current is then set to the value corresponding to the lowest point desired

and reversed several times until successive readings of the induction are in agreement. The specimen is then in a cyclic condition, and the observed induction is the normal induction. The magnetizing force is then determined either by taking a ballistic deflection or by measuring the magnetizing current according to the type of apparatus being used. Additional points on the normal-induction curve are determined in the same way, except that repeated demagnetization is not required if each point so determined is higher than any preceding one. It is the practice of some observers to start with the highest point to be determined and demagnetize from each point to the next lower one. This is sometimes the preferable procedure, especially if otherwise the specimen is likely to be heated unduly.

If points on a hysteresis loop are to be determined, cyclic condition is first obtained by reversals of the magnetizing current corresponding to the tip of the loop and corresponding values of B and H observed. The current is then suddenly reduced in value by opening the switch C' , thus inserting into the circuit R' . For points on the negative side of the H axis, switches C and C' are operated simultaneously, thus reversing as well as reducing the magnetizing force. The observed values of the corresponding changes in B and H are subtracted from the values at the tip, and the results thus obtained are taken as the coordinates of the required point on the hysteresis loop. The process is repeated for other points, care being taken to reestablish cyclic condition at the tip before each reading.

In setting up apparatus for ballistic tests, it is important to twist the conductors in both primary and secondary circuits and to locate the mutual inductor with reference to the rest of the apparatus so that errors due to stray fields will be negligible.

TESTS WITH ALTERNATING CURRENTS

Magnetic tests with alternating currents for the measurement of core loss, a - c permeability, and incremental permeability are ordinarily made in accordance with the specifications of the ASTM which should be consulted for details.² Such tests are made on flat-rolled materials cut into strips of appropriate size. The strips are assembled in four bundles in the form of a square within a set of coils generally known as a "test frame." The dimensions of the strips and of the test frame and the form of joint at the corners of the magnetic circuit depend upon the type of test. Three different types of test follow:

54. 50-cm Epstein Test for Core Loss. The test specimen for this core-loss test consists of strips 50 cm ($19\frac{1}{8}$ in.) long and 3 cm ($1\frac{3}{8}$ in.) wide. The standard sample is composed of 10 kg (22 lb) of strips, but 5 kg (11 lb) may be used. For the usual type of material, half of the strips are cut parallel and half at right angles to the direction of rolling. Under certain conditions, however, material manufactured in strip form or materials having pronounced directional properties may be tested with strips all cut in the same direction. The strips are assembled in four equal bundles and inserted in the test frame so as to form a square with butt joints. The bundles are ordinarily arranged so that opposite sides of the square consist of material cut in the same direction. The strips are held in place at the corners by clamps.

The four coil forms of the test frame are uniformly wound with two sets of coils connected in series to form primary and secondary windings of 600 turns each. The connections are as shown in Fig. 19, except that for the core-loss test the mutual inductor is not used and need not be connected.

An emf of approximately sinusoidal wave form is applied to the primary winding, and a voltmeter and the potential coil of a wattmeter are connected to the secondary winding. The wattmeter, whose current coil is connected in series with the primary winding, should be especially designed for low power factors. The impressed voltage is adjusted, preferably by means of a suitable autotransformer, until the voltage of the secondary is that given by the equation

$$E = 4\pi \frac{NfBM}{4lD10^3}$$

where ff = form factor* of primary emf = 1.11 for sine wave
 N = number of secondary turns = 600
 f = frequency, cycles = 60
 B = maximum induction, kilogausses
 M = total mass, kg
 l = length of strips, cm = 50
 D = density, g per cu cm

Standard tests are made at specified values of maximum induction. For tests at 15 kilogausses or higher, or whenever the form factor of the applied emf departs from the value 1.11 by more than 1 per cent, a voltmeter reading average volts is used in parallel with the rms voltmeter. The scale of such an instrument is conveniently marked in terms of the average volts times 1.11, in which case the voltage, as calculated for a sine wave, is held on the average voltmeter.

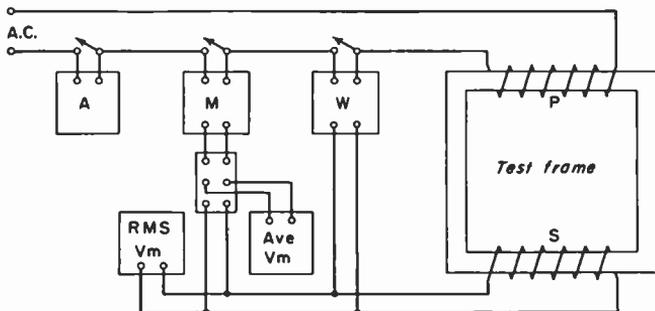


FIG. 19. Circuit for core-loss and a-c permeability tests.

When the voltage and the frequency have been adjusted to the proper values, the wattmeter indicates the total loss, including the loss in the secondary circuit. The loss in the secondary circuit can be calculated in terms of the rms voltage and the resistance. Subtracting this correction from the total loss gives the net loss in the steel, and dividing this value by the mass in kilograms gives the core loss in watts per kilogram.

55. 25-cm Epstein Test for Core Loss and A-c Permeability. In recent years, the manufacturers of electrical sheet have been able to improve the uniformity of the magnetic properties of the various commercial materials to such an extent that it is now possible to obtain a sufficiently representative sample of a given heat or "lift" of steel by selecting not over 2 kg of strips instead of the 10 kg previously required. The 25-cm Epstein test is designed to test specimens weighing from 0.5 to 2 kg, but for general commercial testing the 2-kg sample is the standard. On account of the shorter length of the magnetic circuit, a better joint at the corners is required than the butt joint employed in the 50-cm test.¹¹ A double lap joint has been found to give best results and consequently a minimum length of 28 cm is needed. However, longer strips can be used up to a maximum length of 50 cm. The width of the strips is 3 cm.

The connections and testing procedure for measuring core loss are essentially the same as for the 50-cm test (Fig. 19) but, on account of the smaller cross-sectional area, the number of turns in the test frame windings is increased to 700 and somewhat more sensitive instruments are required. Also, in calculating the core loss it is necessary to take into account the additional material at the corners. This is done by using the "active weight" calculated on the assumption that the effective length of the magnetic circuit is 94 cm. This value was determined by experiment.

For the measurement of magnetizing force when determining a-c permeability the

* Form factor is the ratio of the effective (rms) value of the emf to its half-period average value.

mutual inductor M (Fig. 19) is used. The secondary voltage measured with the average voltmeter is proportional to the crest value of the magnetizing current. The magnetizing force is calculated from the equation

$$H = \frac{0.4\pi NI_c}{l_2} = 10I_c$$

where H = magnetizing force, oersteds
 N = number of turns = 700
 I_c = crest current, amp
 l_2 = assumed length of specimen = 88 cm

The value of B in gauss is determined from the secondary voltage as for the core-loss test and the permeability, μ , is $\mu = B/H$.

56. Normal and Incremental Permeability and Core Loss at Low Inductions. The methods described above are not sufficiently sensitive for testing laminated-core materials at the low inductions usually employed in apparatus used in many types of communications equipment. Also, these methods do not provide for testing under the condition in which a unidirectional magnetizing force is applied in addition to the alternating magnetizing force. For testing at low alternating inductions, either with or without a superimposed unidirectional magnetizing force, two different methods are specified by the ASTM² (1) an a-c bridge method or (2) an a-c potentiometer method. The tests are made at frequencies of either 60 or 1,000 cycles. The standard induction with no superimposed unidirectional magnetizing force is either 10 or 1,000 gauss at 60 cycles or 10 gauss at 1,000 cycles. With unidirectional magnetizing force applied, the standard induction is 10 gauss at either frequency. Values of unidirectional magnetizing force up to 2 oersteds are employed. The test frame is similar to that for the 25-cm Epstein test. Three windings are provided, an inner winding of 100 turns, an intermediate winding of 1,000 turns, and an outer winding of 100 turns, which is used for the application of the d-c component of magnetizing force. The specimens are cut and assembled in the same way as for the 25-cm Epstein test described above. The test frame and specimen are the same for either the bridge or potentiometer method.

Figure 20 is a diagram of connections for the bridge method. The a-c supply should be of approximately a sine wave form having not more than 10 per cent of total harmonics and should be effectively insulated from the bridge circuit by a suitable coupling transformer as shown. A storage battery provides a steady source of d.c. As shown in the diagram, one winding of the test frame constitutes one arm of the bridge. For the 60-cycle test a vibration galvanometer is recommended; for the 1,000-cycle test a telephone receiver may be used. The detector may be preceded by a suitable amplifier.

Before the test is made, the specimen is thoroughly demagnetized. In view of the fact that the permeability of most materials at low inductions drifts to an appreciable

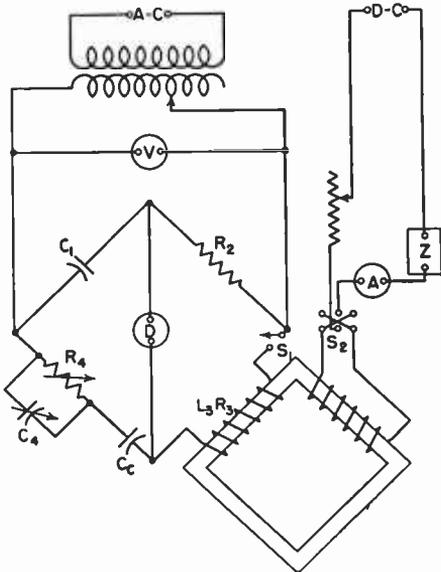


FIG. 20. Bridge method for a-c tests at low inductions.

extent with time after demagnetization,¹² this should be done at least 24 hr previous to making the test. During the interval between demagnetization and testing, the specimen should be protected against stray magnetic fields or mechanical vibration.

The constants of the bridge circuit depend upon the kind of material to be tested. For the 60-cycle test on materials that have exceptionally high permeability and low loss, such as cold-reduced silicon steel or nickel-iron alloys, the inner 100-turn winding of the test frame is used and R_2 is 10 ohms. For the 60-cycle test on ordinary silicon steel, or materials having similar magnetic properties, the 1,000-turn coil is used and R_2 is 100 ohms. In either case C_1 is 1 μ f. C_c balances the ohmic resistance of the test-frame coil and has a value

$$C_c = \frac{R_2 C_1}{R_3}$$

where R_3 is the ohmic resistance of the test-frame winding, R_4 is nonreactive and has a range of 10,000 ohms in steps of 1 ohm, C_4 has a range up to 2 μ f in steps of 0.001 μ f.

Measurements of incremental permeability or core loss are always made in the order of increasing values of unidirectional magnetizing force. Direct current is supplied to the outer 100-turn winding through the reactor Z , which should have an inductance of 1 h or more when carrying current equivalent to a magnetizing force of 2 oersteds. This reactor is for the purpose of limiting the a-c current in the d-c circuit to a negligible value. The d-c current is adjusted to the required value with S_1 open and reversed by means of S_2 several times to establish a cyclic condition in the specimen. S_1 is then closed and the test made. Values of normal permeability or core loss are made with the d-c circuit open.

The d-c magnetizing force is calculated from the equation

$$H_{dc} = \frac{0.4\pi N I_{dc}}{l}$$

where H_{dc} = unidirectional magnetizing force, oersteds

N = number of turns

I_{dc} = direct current, amp

l = assumed length of the magnetic circuit

When $N = 100$ and $l = 94$, this reduces to

$$H_{dc} = 1.34 I_{dc} \quad \text{or} \quad I_{dc} = 0.748 H_{dc}$$

In making a test either with or without unidirectional magnetizing force, the voltage across the bridge is set according to the equation*

$$E = 0.0707 AB$$

where A = cross-sectional area of the specimen, cm

B = maximum induction, gausses

R_4 and C_4 are adjusted to balance the bridge. When the bridge is balanced, the permeability is calculated from the equation

$$\mu \text{ (or } \mu_{\Delta}) = \frac{0.748 R_4}{A} \times \frac{R_2}{N^2} \times 10^4$$

where μ = normal permeability

μ_{Δ} = incremental permeability

R_4 = resistance, ohms

A = cross section, sq cm

R_2 = resistance, ohms

N = number of turns

The loss is calculated from the equation

$$P_c \text{ (or } P_{\Delta}) = E^2 C_4 \omega^2 C_1 R_2$$

* Under certain conditions it may be necessary to make a correction to the voltage thus calculated. The ASTM specifications² should be consulted for details.

- where P_c = total loss in watts without unidirectional magnetizing force
- P_Δ = total loss in watts with unidirectional magnetizing force
- E = emf, volts
- C_1 = capacitance, farads
- ω = 2π times the frequency, cycles
- C_2 = capacitance, farads
- R_2 = resistance, ohms

The total core loss in watts divided by 84 per cent of the mass of the specimen in kilograms is the value of core loss in watts per kilogram.

Figure 21 is a diagram of connections for the a-c potentiometer method. The a-c potentiometer is of the coordinate type, which indicates voltage in terms of two components having a quadrature phase relation. A phase-shifting device (not shown) is used to adjust the phase of either the potentiometer current or the magnetizing current. The d-c circuit and procedure for applying a unidirectional magnetizing force are the same as for the bridge method.

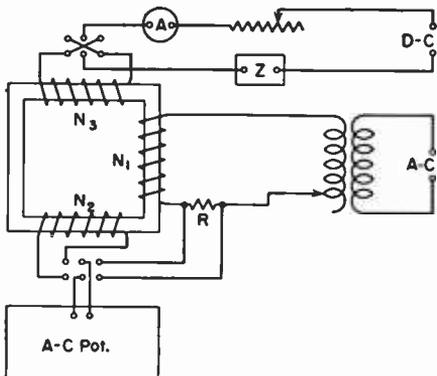


FIG. 21. Alternating-current potentiometer method for tests at low inductions.

In making a test, the 100-turn coil is used as the primary winding and the 1,000-turn coil is the secondary unless the secondary voltage exceeds the range of the potentiometer, in which case the two windings must be interchanged. The potentiometer is connected to the secondary coil, and the magnetizing current and phase relations are adjusted so that the total voltage is read on the in-phase dial of the potentiometer, the other dial being set at zero. The voltage corresponding to a given induction is calculated from the equation

$$E = \sqrt{2} \pi f N_2 A B \times 10^{-8}$$

- where E = emf, volts
- f = frequency, cycles
- N_2 = number of turns in the secondary winding
- A = cross-sectional area, sq cm
- B = induction, gausses

With the secondary voltage set at the proper value, the two components of the magnetizing current are measured by observing the drop across the noninductive resistor R in series with the magnetizing winding. The power component is in phase with the voltage and the magnetizing component is in quadrature. Permeability and core loss are calculated in terms of secondary voltage and the quadrature and in-phase components of the current respectively, by using the equations

$$\mu \text{ (or } \mu_\Delta) = 52.9 \frac{B}{N_1 I_q} \quad P_c \text{ (or } P_\Delta) = \frac{N_1}{N_2} E I_p$$

- where μ = normal permeability
- μ_Δ = incremental permeability
- B = induction, gausses
- N_1 = number of primary turns
- I_q = quadrature component of the current, amp
- P_c = total loss in watts without unidirectional magnetizing force
- P_Δ = total loss in watts with unidirectional magnetizing force
- N_2 = number of secondary turns
- E = secondary emf, volts
- I_p = in-phase component of the current, amp

The loss in watts per kilogram is obtained by dividing the total loss by 84 per cent of the mass of the specimen in kilograms.

Recently, several methods have been devised by which hysteresis loops are traced on the screen of a cathode-ray oscilloscope. These methods^{13,14} provide a rapid and convenient means of inspecting magnetic materials. The precision of such tests, however, is somewhat less than is attained with permeameters or the a-c methods described above.

57. Tests of Permanent Magnets. The testing of permanent-magnet materials for determining their demagnetization curves is pretty well standardized⁹ and need not be discussed here. However, there are some principles involved in the testing of finished magnets that merit some consideration.

It is a quite common practice to judge the strength of a magnet roughly by the force of attraction between it and a piece of soft iron. If the purpose for which the magnet is to be used is to attract or hold an armature, this is a logical method of test, and the attractive force may be measured quantitatively for the purpose of comparing magnets of the same size and shape. However, if the magnet is to be used for any

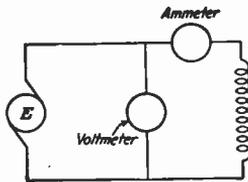


FIG. 22. Circuit for measurement of inductance.

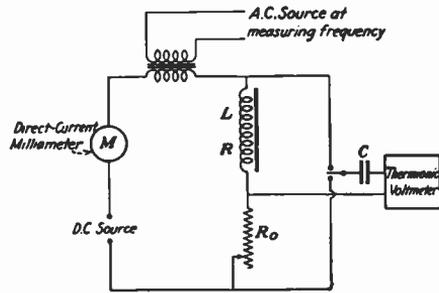


FIG. 23. Measurement of iron-core coil carrying a.c. and d.c.

other purpose, the application of such a test may not only give unsatisfactory results but also may prove to be detrimental to the magnet. The characterization of a permanent-magnet material by stating the weight a magnet can lift as related to its own weight is usually misleading, because this relationship is a function not only of the magnetic quality of the material but also of the size and shape of the magnet and the nature of the contact between the magnet and the mass lifted.

Magnets are customarily tested by either one of two methods. One method consists in determining the flux in the neutral zone by quickly removing a test coil connected to a ballistic galvanometer or fluxmeter. The other method consists in determining the field strength in the gap of special pole pieces to which the magnet is applied. The field strength may be measured in terms of the deflection of a pivoted coil in which a definite current is flowing, the emf induced in an armature rotated at a definite speed, or by a test coil and ballistic galvanometer. A bismuth spiral might also be used, but this is less satisfactory than any of the other methods.

Unless the special pole pieces have the same permeance as the pole pieces to which the magnet is to be applied in use, neither of these methods duplicates the conditions under which the magnet is to function and, consequently, the results must be interpreted with care. If the magnet is designed in such a way as to require magnetization with pole pieces in place, with subsequent magnetic stabilization by partial demagnetization, it will be necessary to remagnetize after testing. Otherwise, the required strength will not be obtained in the finished apparatus. Furthermore, a single test will give no indication of whether or not the magnet has been properly aged. Thermal stabilization must be applied to the finished apparatus after all other treatments have

been applied, and tests as to its effectiveness must be made by temperature tests on the finished apparatus.

58. Measurements of Inductance at Low Frequencies.* The measurement of the inductance of air-core coils at low frequencies is relatively simple, as the inductance is sensibly constant with change in frequency and current. Iron-core inductors do not have a fixed inductance under all conditions, and measurements on them must be made under conditions that duplicate as nearly as possible the conditions under which the inductor is used.

A simple method of approximate measurement uses the circuit of Fig. 22. An a-c voltage of known frequency is applied at E , and the current and voltage read on the meters. The voltmeter reading divided by the ammeter reading gives the impedance and, if the resistance is measured by a d-c-bridge or voltmeter-ammeter method,

$$L = \sqrt{\frac{Z^2 - R^2}{4\pi^2 f^2}} = \frac{0.159}{f} \sqrt{Z^2 - R^2}$$

The method is usable for iron-core coils that carry a.c. only, provided the measuring current is adjusted to the value that the coil carries in use. If measurements are made at a number of current values, the curve of inductance against current can be plotted. The results obtained by this method are generally slightly larger than the true values of inductance because the a-c resistance, particularly in iron-core coils, is greater than the d-c resistance.

59. Measurement of Inductance of Iron-core Coils. When an iron-core coil must carry relatively large d.c. upon which is superimposed a small value of a.c., its inductance is dependent upon the magnitudes of the two currents flowing through it, and other methods must be used.

The impedance of an iron-core coil carrying d.c. and a.c. can be measured by the circuit of Fig. 23. The d.c. through the circuit is adjusted to the value carried by the coil during operation, and the a-c source adjusted to impress a voltage across the coil (measured by the thermionic voltmeter) equal to the a-c voltage across it under operating conditions. The resistance R_0

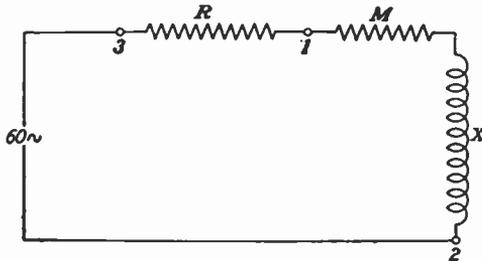


FIG. 24. Circuit for determining inductance of iron-core coil.

is then varied until the alternating voltage across it is equal to that across the coil, as measured by the thermionic voltmeter. Then the impedance of the coil at the measuring frequency is equal to R_0 . Readjustments of the impressed direct and alternating voltages may be necessary as R_0 is changed. The capacitor C prevents the direct voltages across the coil and resistor from affecting the thermionic voltmeter. From the impedance and the resistance of the coil, the inductance may be calculated by the equation above.

In Fig. 24 is a simple method of arriving at the impedance of an iron-core coil based on the supposition that the inductance is high compared to the resistance. The voltage across R and X is measured with a vacuum-tube voltmeter, for example. Then $E_r/R = I$ and $E_x/I = X = (E_x/E_r) \times R$, whence

$$X = \frac{R}{E_r} E_x$$

In the general case in which M represents the total losses of the coil, the power factor of the inductance is $\cos \theta$ and

* The following three sections are included from previous editions of this handbook.

$$\cos \theta = \frac{E_{2-3} - E_1 - E_r}{2E_1E_r}$$

and the total losses in the core and winding may be thus obtained.

Once the impedance, reactance, and inductance of a coil have been determined, the permeability and finally the magnetizing force and flux density of an iron-core coil may be obtained. Thus the a-c flux density

$$B_{\max} = \frac{E_{\text{eff}} \times 10^8}{4.44 \times f \times N \times A \times K} \text{ gaussess}$$

where E_{eff} = rms voltage across the coil

f = frequency, cps

N = number of turns in the winding

A = cross section of the core, sq cm

K = core-stacking factor

The polarizing mmf resulting from the d.c. in the winding, in gilberts per centimeter, is given by

$$H_0 = \frac{1.256NI}{l}$$

where N = number of turns in the winding

I = d.c., amp

l = length of magnetic circuit, cm

To get mmf in ampere-turns per inch, multiply H_0 by 2.032.

The following table (Allegheny-Ludlum Steel Company) gives values of B_{\max} and H_0 found in practice.

Coil	B_{\max} , gaussess	H_0 , gilberts/cm
Detector-stage audio transformer.....	0.5-10	0.6-1.2
Second-stage a-f transformer.....	250	1.5
Push-pull output transformer with two primaries.....	7,000	0
Polarized output transformer.....	4,200	6.7
Heavy-duty filter reactor (80 ma).....	300	27

60. Turner Constant-impedance Method. For measurements involving a.c. only, the constant-impedance method (of Turner*), shown in Fig. 25, is used. The method

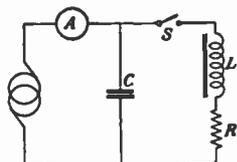


FIG. 25. Turner constant-impedance method.

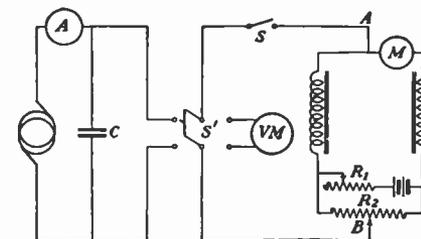


FIG. 26. Measuring circuit for coils carrying a.c. and d.c.

is based upon the fact that, when $1 - \omega^2LC = 0$, the impedance of the parallel circuit is equal to ωC and is independent of the resistance in the inductive branch. Consequently the line current will have the same magnitude with the switch open or closed.

* Turner, H. M., Constant Impedance Method for Measuring Inductance of Choke Coils, *Proc. IRE*, 16, 1559, 1928.

To measure any value of inductance, then, it is only necessary to adjust the capacitance so that the reading of the ammeter A is the same for both positions of the switch. Then

$$L = \frac{1}{2\omega^2 C}$$

When the coil must carry d.c. as well as a.c., the circuit of Fig. 26 may be used for the inductance measurement. Two similar inductors are used, the d.c. through them being adjusted to the proper value by means of the resistor R_1 and measured by means of the d-c ammeter M . The switch S' is then thrown to the right and the resistor R_2 adjusted to make the constant-potential difference between the points A and B zero. Then, with S' thrown to the left, the inductance measurement may be carried out in the manner already described. The result is the inductance of the two coils in parallel, which is one-half the inductance of one coil.

61. Applications. The principal applications involving the use of magnetic materials in radio or other electronic devices are (1) electromagnets, (2) transformers and inductors, (3) permanent magnets.

1. The principal application of electromagnets is in relays.^{15,16,17} Unless a relay is to be operated with a.c., the core material is usually of solid material. If a.c. is used, the core must be laminated to minimize the demagnetizing effect of induced eddy currents. There are many types of relay in use, and the design required depends upon the functions and conditions of use. If the relay must operate with a minimum of power and can operate at a moderate flux density, some of the nickel-iron alloys may be best to use for core and armature material. However, if higher flux densities can be employed to advantage, a low silicon-iron alloy known as "relay iron" is often used. The silicon tends to decrease aging and the tendency to "stick" after prolonged operation.

Loud-speaker magnets constitute another application of electromagnets. The voice coil of a loud-speaker operates in the air gap of an electromagnet energized by a constant d.c. Solid core materials are used in such electromagnets. In both relays and loud-speaker magnets, the aim of the designer is to keep as much of the total magnetic flux as possible in the air gap where the work is done and so design the magnetic structure that there will be a minimum of stray or leakage flux in other parts of the magnetic circuit. Since there is no insulator of magnetic flux, this involves proper shaping of the parts of the magnetic circuit and location of the magnetizing winding.

Another type of electromagnet that is coming into use is the recording, playback, and erasing heads used in magnetic recording. The magnetic core of such a head is usually constructed of one of the high-permeability nickel-iron alloys. The most common forms utilize the fringing flux across the very short air gap (of the order of 0.001 in.) for recording and erasing. A high value of initial permeability is particularly important for playback heads.

2. The core material for power transformers¹⁸ is usually one of the various grades of silicon steel, the particular grade selected depending on the type of service, whether continuous or intermittent. If a minimum of distortion is required as in interstage coupling transformers, it is necessary either to make the core of relatively large cross section if silicon steel is used or to employ one of the high-permeability nickel-iron alloys. Considerations of space and cost usually govern the choice.

The inductance of coils having magnetic cores is not constant but depends upon the current flowing. The variation is minimized either by an air gap in the magnetic circuit or by the use of powdered materials. In either case the effective permeability of the core material is considerably reduced, but the inductance is still much greater than could be obtained with the same winding without the magnetic core.

3. Permanent magnets⁹ are used for a great variety of purposes, chiefly in indicating meters, loud-speakers, and polarized relays. The choice of material for a given purpose depends upon a number of factors such as space requirements, degree of stability required, and cost. In many applications where space and weight are not important, the older types of tungsten or chromium magnet steels are satisfactory and more economical in cost than the newer alloys. In applications where space requirements are important and cost is a secondary consideration, the newer alloys such as the Alnicos, Cunife, Cunico, or Vectolite are usually employed. Sometimes the more expensive alloys will have to be used because no other materials are capable of giving the required performance.

References

1. Bosorth, R. M.: Present Status of Ferromagnetic Theory, *Elec. Eng.*, **54**, 1251, 1935; The Physical Basis of Ferromagnetism, *Bell System Tech. J.*, **19**, 1, 1940.
2. "Metals," Part I, latest ed., ASTM Standards.
3. Sanford, R. L.: Permanent Magnets, *Nat. Bur. Standards Circ. C448*, 1944.
4. Ewing, J. A.: "Magnetic Induction in Iron and Other Metals," 3d ed., p. 138, Electrician Printing Publishing Co., Ltd., London, 1900.
5. Burrows, C. W.: The Determination of the Magnetic Induction in Straight Bars, *Nat. Bur. Standards Bull. S117*, **6**, 31, 1909.
6. Fahy, F. P.: A Permeameter for General Magnetic Analysis, *Chem. & Met. Eng.*, **19**, 339, 1918.
7. Sanford, R. L., and E. G. Bennett: An Apparatus for Magnetic Testing at Magnetizing Forces up to 5,000 Oersteds, *J. Research Nat. Bur. Standards*, **23**, 415, 1939.
8. Babbitt, B. J.: An Improved Permeameter for Testing Magnet Steel, *J. Optical Soc. Am.*, **17**, 47, 1928.
9. Sanford, R. L.: Magnetic Testing, *Nat. Bur. Standards Circ. C456*, 1946.
10. Gokhale, S. L.: The Saturation Permeameter, *J. AIEE*, **47**, 196, 1928.
11. Burgwin, S. L.: Measurement of Core Loss and A-c Permeability with the 25-cm Epstein Frame, *Proc. ASTM*, **41**, 779, 1941.
12. Sanford, R. L.: Drift of Magnetic Permeability at Low Inductions after Demagnetization, *J. Research Nat. Bur. Standards*, **13**, 371, 1934. RP714.
13. Wiegand, D. E., and W. W. Hansen: A 60-cycle Hysteresis Loop Tracer for Small Samples of Low-permeability Material, *Trans., AIEE*, **66**, 1947.
14. Zemsky, Joseph: Quantitative Determination of Magnetic Properties by Use of Cathode-ray Oscilloscope, *AIEE Tech. Paper* 47-119, April, 1947.
15. Packard, C. A.: "Relay Engineering," Struthers-Dunn, Inc., Philadelphia, 1945.
16. Fisher, R. T.: Graphic Solution of Design Problems Involving Sensitive Relays, *Electronics*, October, 1943.
17. Schoemann, A. F.: Selecting the Right Relay, *Elec. Mfg.*, June, 1936.
18. Massachusetts Institute of Technology, "Magnetic Circuits and Transformers," Wiley, 1943.
19. Salt peter, J. L.: Permanent Magnets, *Proc. IRE (Australia)*, May, June, 1947.
20. Underhill, Earl M.: Permanent Magnet Design, *Electronics*, December, 1943; January, February, April, 1944.

CHAPTER 4

CAPACITANCE

BY E. L. HALL¹

1. Capacitance. Capacitance is one of the three electrical quantities present in all radio circuits. The radio engineer endeavors to concentrate capacitance in definite well-known forms at definite points in the circuits, but capacitance exists between conductors in the circuits at different potentials and between the various conductors and the ground. Such capacitances, usually small, are ordinarily of no importance in the case of l-f or a-f currents but may be of great consequence in r-f circuits.

A *condenser* or *capacitor* is an electrical device in which capacitance plays the main role. Although some inductance and some resistance are present, these quantities are often of minor importance and sometimes may be neglected.

A capacitor has three essential parts, two of which are usually metal plates separated or insulated by the third part called the *dielectric*.

The amount of electricity that the capacitor will hold depends on the voltage applied to it. This may be expressed as $Q = C \times V$. The *capacitance* is the ratio of the quantity of electricity and the potential difference or voltage, or $C = Q/V$ where Q is given in coulombs, C in farads, and V in volts. The capacitance is dependent on the size and spacing of the plates and the kind of dielectric between them.

2. Units of Capacitance. The unit of capacitance is the *farad*. A capacitor has a capacitance of one farad when one coulomb of electricity can be added to it by an applied voltage of one volt. As this unit is too large for practical use, a smaller unit, the microfarad, abbreviated μf , or one-millionth of a farad, is used. Condensers for radio circuits usually have capacitances between a few thousandths and a few millionths of a microfarad. Another unit, the micromicrofarad, is often used. It is abbreviated $\mu\mu f$.

Another unit of capacitance sometimes used is the *centimeter*. The centimeter is equal to 1.1124 $\mu\mu f$.

3. Electrical Energy of Charged Capacitor. Work is done in charging a capacitor because the dielectric opposes the setting up of the electric strain or displacement of the electric field in the dielectric. The energy of the charging source is stored up as electrostatic energy in the dielectric.

The work done in placing a charge in the capacitor is

$$W = \frac{1}{2} Q \times V = \frac{1}{2} CV^2 = \frac{Q^2}{2C}$$

where W is expressed in joules, Q in coulombs, and V in volts. The work done in charging the capacitor is independent of the time taken to charge it.

4. Power Required to Charge Capacitor. The average power required to charge a capacitor is given by the equation

$$P = \frac{1}{2} \frac{CV^2}{t}$$

where P is expressed in watts, C in microfarads, V in volts, and t in seconds. If the capacitor is charged and discharged N times per second, the above equation becomes

$$P = \frac{1}{2} CV^2N$$

¹Radio Engineer, High Frequency Standards Section, Central Radio Propagation Laboratory, National Bureau of Standards.

If an alternating emf of frequency f is used in charging the capacitor, the equation may be written

$$P = CE_0^2f$$

where P = power, watts

C = capacitance, farads

E_0 = maximum value of voltage

f = frequency, cps

5. Dielectric Materials. The dielectric of a capacitor is one of the three essential parts. It may be found in solid, liquid, or gaseous form or in combinations of these forms in a given capacitor.

The simplest form of capacitor consists of two electrodes or plates separated by air. This represents a capacitor having a gaseous dielectric. If this imaginary capacitor has the air between the plates replaced by a nonconducting liquid, such as transformer oil, and if the distance between the plates is the same as in the first case, it would be found that the capacitance was increased several times because the oil has a higher value of dielectric constant than air which is usually taken as 1.

If the space between the plates is occupied by a solid insulator, a capacitor would result, which would be practical as far as the possibility of constructing it is concerned. It would be found, in this case also, that the capacitance of the capacitor was several times larger than when air was the dielectric.

The mechanical construction of either air or liquid dielectric capacitors requires the use of a certain amount of solid dielectric for holding the two sets of plates.

There are a great many dielectric or insulating materials available from which the engineer may choose. It is often found that a material that is very good from the electrical standpoint is poor mechanically, or vice versa. Air is the gas generally used as a dielectric. Compressed air has been used in some high-voltage capacitors, and compressed nitrogen and carbon dioxide are also in use.

Several kinds of oil have been used in capacitors, such as castor oil, cottonseed oil, and transformer oil. Electrolytic capacitors have come into use in radio equipment for use as filters and by-pass capacitors where a large capacitance in compact form is required and either a d.c. or pulsating d.c. is applied.

Among the solids used as the dielectric are mica, ceramic materials, and paper. Solid insulators used as mechanical supports include quartz, glass, Isolantite, porcelain, mica, amber, hard rubber, Victron, and Mycalex.

6. Dielectric Properties of Insulating Materials. Such properties as surface and volume resistivity, dielectric strength or puncture voltage, dielectric constant, and absorption are often considered in d-c and commercial-frequency applications. Such data are of little value if the insulating material is to be used at radio frequencies. For the latter application r-f measurements of various properties of the material are essential. A material that may be a satisfactory insulator for low frequencies may be worthless as an insulator at radio frequencies.

One of the most important properties of an insulator for radio frequencies is its power loss. This includes several factors which are difficult to separate but together indicate its suitability for radio purposes. The general idea of the imperfection of a capacitor is brought out in several names such as "power loss," "power factor," and "phase difference," but they are not identical terms.

Dielectric constant is another important property of a material that has a definite bearing upon its use at radio frequencies.

Neither power loss nor dielectric constant alone can be used in selecting the best insulator for a particular application at radio frequencies.

7. Dielectric Constant. The *dielectric constant* K of an insulating material is the ratio of the capacitance C_x of a condenser using the material as the dielectric, to the capacitance C_a using air as the dielectric, or $K = C_x/C_a$. This property of the material is sometimes called *inductivity* or *specific inductive capacity*.

The dielectric constant of a material is not a constant in the true sense of the word,

but varies with the frequency, moisture content, temperature, voltage applied, and manner of applying it.

Table 1 gives the dielectric constants of a large number of electrical insulating materials.

8. Power Loss, Phase Difference, Power Factor, and Loss Factor. Electrical insulating materials are not perfect in their insulating qualities, and there is a certain amount of power absorbed in them when used in an a-c circuit. A measurement of the power loss is the best single property that gives an indication of the suitability of an insulating material for use in radio circuits. Power loss can be expressed by a number of quantities, the most commonly used being resistance, power factor, phase difference, and phase angle.

When a.c. flows in a capacitor, the voltage across the capacitor lags somewhat less than 90 deg behind the current as shown by the angle θ (Fig. 1), called the *phase angle*. The complement of the phase angle, is called the *phase difference*. The cosine of the phase angle is called the *power factor*. The power loss in the insulating material is

$$P = EI \cos \theta \quad \text{or} \quad P = EI \sin \psi$$

where E = voltage across the capacitor

I = current through the capacitor, amp

θ plus ψ = 90 deg, as shown in Fig. 1.

From the above, $\sin \psi = \cos \theta$, or the sine of the phase difference is equal to the power factor.

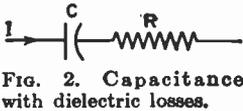


Fig. 2. Capacitance with dielectric losses.

When considering a capacitor having dielectric losses, such as current leakage, brush discharge or corona, dielectric absorption or resistance in the plates, joints, contacts, and leads, it is customary to think of it as a perfect capacitor C with a resistance R in series as shown in Fig. 2.

The voltage vectors may be shown as in Fig. 3, where the resultant voltage E flowing in the circuit is obtained by completing the vector diagram. The angle ψ is quite small for materials suitable for r-f insulators. For small angles the angle $\psi = \tan \psi$. In Fig. 3,

$$\tan \psi = \frac{RI}{I/\omega C} = R\omega C = 2\pi fRC$$

If the resistance, capacitance, and frequency can be measured, the phase difference can be calculated from

$$\psi = 2\pi fRC$$

where ψ = phase difference, radians

f = frequency, cps

R = resistance, ohms

C = capacitance, farads

The following equation is sometimes convenient when wavelength in meters is given

$$\psi = 0.1079 \frac{RC}{\lambda}$$

where ψ = phase difference, deg

R = resistance, ohms

C = capacitance, $\mu\mu\text{f}$

λ = wavelength, m

For small angles, phase difference in radians is equal to power factor (nearly).

Power factor in per cent is 1.745 times phase difference in degrees. Power factor in per cent is given by the following equation:

$$\cos \theta = 2\pi fRC \times 10^{-7}$$

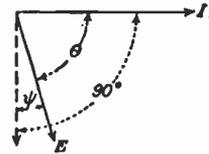


FIG. 1. Phase in a capacitive circuit.

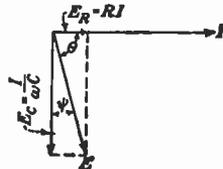


FIG. 3. Vector relations in a capacitor with dielectric losses.

where $\cos \theta$ = power factor, per cent
 f = frequency, kc
 R = resistance, ohms
 C = capacitance, μmf

The leakage of electricity by conduction through the dielectric or along its surface contributes to the phase difference but is generally negligible at high frequencies. A capacitor having leakage may be represented by a perfect unit capacitor with a resistance in parallel as shown in Fig. 4. The current divides between the capacitance and the resistance, I_R through the resistance being in phase with the applied voltage E , and I_C through the capacitance leading E by 90 deg as shown in Fig. 5. The resultant current I leads E by (90 deg - ψ), where ψ is the phase difference. In Fig. 5,

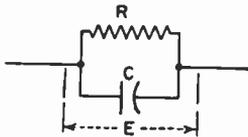


FIG. 4. Equivalent of capacitance with leakage.

$$\tan \psi = \frac{E/R}{\omega CE} = \frac{1}{\omega RC} \quad \text{or} \quad \psi = \frac{1}{\omega RC}$$

Power factor is a term that involves all the power losses in a capacitor. If the total power loss in a condenser is W watts, the voltage applied to it is V volts (rms), and the current flowing through it is I amp (rms), the power factor of the unit is W/VI . The relation between I (amp) and V (volts) for a unit of capacitance C (μf) operating at a frequency f is

$$I = \frac{2\pi f CV}{10^6} = \frac{\omega CV}{10^6}$$

The power factor in per cent may be written

$$\cos \theta = \frac{W \times 10^6}{2\pi f CV^2} = \frac{W \times 10^6}{\omega CV^2}$$

Referring again to Fig. 2 showing the perfect capacitor C and resistance R replacing the actual capacitor, the value of R can be calculated from the equation $W = I^2R$. The quantity R is known as the *equivalent resistance* of the capacitor at the given frequency.

The expression $W \times 10^6/\omega CV^2$ for power factor can be changed into the expression involving resistance, capacitance, and ω by substituting I^2R for W and then substituting $\omega CV/10^6$ for I , giving power factor equal to

$$RC\omega \times 10^{-6}$$

The power loss or rate at which heat is generated within a dielectric is proportional to the product of the dielectric constant and the power factor. This product is known as the "loss factor" of the dielectric. The power loss in watts per cubic inch may be calculated from the following:¹

$$P = 2\pi f G^2 K \tau \times 0.2244 \times 10^{-12}$$

where P = power loss, watts per cu in.

f = frequency, cps

G = voltage gradient in dielectric, rms volts per in.

K = dielectric constant

τ = power factor of dielectric

9. Table 1 gives dielectric constant and power factor at certain frequencies of a large number of electrical insulating materials, as obtained from the sources given at the end of the table. In some cases data from different sources do not agree, but differences in composition, method of making measurements, and condition of samples may account for such disagreements.

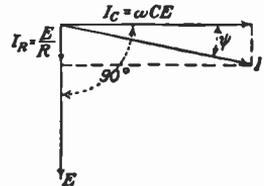


FIG. 5. Vectors in capacitor with leakage.

¹ Hoch, E. T., Power Losses in Insulating Materials, *Bell System Tech. J.*, 1, 110, November, 1922.

Table 1. Values of Dielectric Constant and Power Factor for Electrical Insulating Materials at Radio Frequencies

Material	Frequency, Mc	Dielectric constant	Power factor	Source
Alsimag No. 243, Forsterite.....	1	6.2	0.0004	1,3
	10	6.2	0.0003	1,3
	100	6.1	0.0003	1,3
Alsimag No. 196, Steatite.....	1	5.8	0.0021	1,3
	10	5.7	0.0015	1,3
	100	5.6	0.0014	1,3
Alsimag No. 211, Steatite.....	1	5.8	0.0004	1
	10	5.7	0.0003	1
	60	4.4	0.0004	2
	100	5.7	0.0003	1
	120	4.4	0.0005	2
Alsimag No. 222, Steatite.....	1	5.5	0.0002	1
	10	5.5	0.0002	1
	100	5.5	0.0002	1
Amber.....	0.3	0.0048	4
	0.6	0.0050	4
Aniline formaldehyde resin.....	1.0	0.0051	4
	1	3.5	0.007	6
Casein.....	100	3.4	0.004	6
	1	6.2	0.052	6
Celluloid.....	1	6.2	0.05-0.10	5
Cellulose acetate (plastic).....	1	3.9	0.039	6
	100	3.4	0.039	6
Cellulose acetobutyrate.....	1	3.2	0.017	6
	100	3.0	0.019	6
Dilactene.....	3,000	3.39	0.0038	7
Ethyl cellulose.....	1	3.4	0.028	6
	100	3.2	0.024	6
Glass:				
Borosilicate No. 707.....	60	3.7	0.0012	2
	120	3.7	0.0012	2
Borosilicate.....	18	5.1	0.0059	8
Cobalt.....	0.5	7.3	0.007	9
	1	4.0	0.0008	6
Corning No. 707.....	100	4.0	0.0012	6
	1	5.2	0.0048	6
No. 774.....	100	5.0	0.008	6
No. 790.....	1	3.9	0.0006	6
	100	3.9	0.0006	6
No. 7052.....	1	5.1	0.0024	6
	100	5.1	0.0036	6
Nonex No. 772.....	60	4.2	0.0028	2
	120	4.2	0.0025	2
Plate.....	0.1	0.0077	10
	0.5	0.0066	10
Soda lime No. 008.....	1	0.0062	10
	60	6.1	0.0106	2
Hard rubber.....	120	6.2	0.0103	2
	60	3.1	0.0083	2
Low-loss type.....	120	3.1	0.0084	2
	60	2.9	0.0057	2
Isolantite.....	120	3.0	0.0050	2
	0.25-1.5	6.1	0.0018	1
Italian lavite.....	60	4.7	0.0015	2
	120	4.8	0.0017	2
Lava, grade 1137.....	1	5.6	0.0040	1
	10	5.5	0.0035	1
	100	5.4	0.0030	1
Lucite HM119, methyl methacrylate.....	1	2.6	0.015	6
	100	2.6	0.007	6
Lucite.....	1,333	2.58	0.0090	7
	3,000	2.56	0.0087	7
Lustron.....	1	2.5-2.6	0.0001-0.0004	1
Melamine formaldehyde resin.....	1	4.5	0.08	6
	100	4.5	0.03	6
Mica, clear, muscovite.....	1	6.5-8.0	0.0001-0.0006	11
India, muscovite.....	0.1-1.0	7.9-7.1	0.0002-0.0001	12
U.S., muscovite.....	0.1-1.0	8.7-6.6	0.0004-0.0001	12
Mineral oil.....	1	2.7	0.0008	11
Mycalex.....	0.1	8.0	0.002	13
	1.0	8.0	0.002	11
Mycalex, red.....	1,333	5.91	0.0030	7
	White.....	1,333	5.74	0.0033
Mycalex 364.....	1	7.0	0.0021	6
	100	7.0	0.0022	6

Table 1. Values of Dielectric Constant and Power Factor for Electrical Insulating Materials at Radio Frequencies. (Continued)

Material	Frequency, Mc	Dielectric constant	Power factor	Source
Mykroy 38.....	1	6.68	0.0016	1
	10		0.0015	1
Nylon FM-1.....	1	3.6	0.020	6
	100	3.6	0.018	6
Paraffin oil.....	1	2.2	0.0001	6
	100	2.2	0.0004	6
Phenol formaldehyde resins:				
Cast.....	1	8.0	0.05	6
General-purpose.....	1	4.5	0.014	6
Mineral-filled.....	100	4.0	0.014	6
	1	4.4	0.006	6
	100	4.3	0.012	6
	100	8.0	0.08	6
Phenol furfural resins.....	1	5.0	0.04	6
	100	4.0	0.05	6
Plexiglass, methyl methacrylate.....	1	2.6	0.015	6
	100	2.6	0.007	6
Polyethylene.....	1	2.25	0.0003	6
	100	2.25	0.0003	6
Polyindene.....	1	3.0	0.0004	5
Polyisobutylene MW 100,000.....	1	2.22	0.0003	6
	100	2.22	0.0004	6
Polystyrene.....	60	2.6	0.0005	2
	120	2.5	0.0007	2
Polystyrene MW 80,000.....	1	2.53	0.0002	6
	100	2.52	0.0003	6
Polyvinyl carbazole.....	1	2.95	0.0005	6
	100	2.95	0.0006	6
Polyvinyl chlor-acetate.....	1	2.9	0.014	6
	100	2.8	0.009	6
Polyvinyl chloride.....	1	2.9	0.016	6
	100	2.9	0.008	6
Polyvinylidene chloride-saran.....	1	3.0	0.046	6
	100	2.8	0.014	6
Porcelain, wet-process.....	1	6.5-7.0	0.006-0.008	11
Zircon.....	1	9.2	0.001-0.0014	3
Quartz (fused).....	1	4.1	0.002	11
	1	3.8	0.002	6
Clear.....	100	3.8	0.0002	6
Milky.....	60	3.8	0.0003	2
	120	3.8	0.0005	2
	60	3.5	0.0003	2
	120	3.5	0.0005	2
Shellac.....	1	6.0	0.070	5
	1	3.5	0.031	6
	100	3.1	0.030	6
Styraloy 22.....	1	2.4	0.0012	6
	100	2.4	0.0043	6
Styramic.....	1,333	2.4	0.0047	7
	1	2.75	0.0002	6
	100	2.73	0.0002	6
Styramic HT.....	1	2.64	0.0002	6
	100	2.62	0.0002	6
Styramic E 1689.....	3,000	2.55	0.00087	7
Tenite II.....	3,000	2.95	0.031	7
Textolite No. 1422.....	3,000	2.4-2.5	0.0006-0.0009	1
Urea formaldehyde resins.....	1	5.6	0.028	6
	100	5.0	0.05	6
Urea resin, wood-flour filler.....	1	5.7	0.03	5
	10	5.5	0.038	5
	35	5.3	0.042	5
Vinylite V.....	1,333	2.78	0.0076	7
	3,000	2.61	0.0068	7
Wax, beeswax.....	1	2.9	0.025	11
Boler.....	1,333	2.17	0.00019	7
Ceresin.....	1	2.5-2.6	0.0012-0.0021	11
Halowax.....	1	3.7	0.0014	6
	10	3.4	0.105	6
Paraffin.....	1	2.5	0.0097	11
	1	2.25	0.0002	6
	10	2.25	0.0002	6
	1,333	2.17	0.00019	7
Superla.....	1,333	2.26	0.00019	7
	3,000	2.26	0.00015	7

Sources for Tables

1. Manufacturers' bulletins.
2. Miller, J. M., and B. Salzberg: Measurements of Admittances at Ultra-high Frequencies, *RCA Rev.*, **3**, 480-504, April, 1939.
3. Thurnauer, H.: Low-loss Ceramic Dielectric, *Tele-Tech*, **6**, 86, February, 1947.
4. Schott, Erich: Hochfrequenzverluste von Gläsern und einigen anderen Dielektrics, *Jahrb. drahtlosen Tele. Tele.*, **18**, 82-122, August, 1921.
5. Bloomfield, G. F.: Insulating Materials for the Higher Frequencies. *T & R Bull.*, **14**, 635-639, May, 1939; *Radio Tech. Digest*, September-October, 1939, pp. 23-32.
6. "Reference Data for Radio Engineers," 2d ed., Federal Telephone and Radio Corporation, 1946.
7. Englund, C. R.: Dielectric Constants and Power Factors at Centimeter Wave-lengths, *Bell System Tech. J.*, **23**, 114-129, January, 1944.
8. Chaffee, J. G.: The Determination of Dielectric Properties at Very High Frequencies, *Proc. IRE*, **22**, 1020, August, 1934.
9. Hoch, E. T.: Power Losses in Insulating Materials, *Bell System Tech. J.*, **1**, November, 1922.
10. MacLeod, H. J.: Power Losses in Dielectrics, *Phys. Rev.*, **21**, 53-73, 1923.
11. General Electric Company.
12. Lewis, A. B., E. L. Hall, and F. R. Caldwell: Some Electrical Properties of Foreign and Domestic Micas and the Effect of Elevated Temperatures on Micas, *Bur. Standards, J. Research*, **1**, 409, August, 1931.

10. Dielectric Strength. The *dielectric strength* of an insulating material is the minimum value of electric field intensity required to rupture it. Dielectric strength is usually expressed in kilovolts per centimeter of dielectric thickness. The fall in insulation resistance with rise in temperature is a factor of great importance in connection with the breakdown of a dielectric under the applied voltage. Insulating materials are not strictly homogeneous. The current leakage through an insulating material may perhaps be concentrated in a few small paths through the material, and the energy loss due to the leakage, although small, may be large compared with the area through which it is flowing. The paths of the current flowing through the dielectric become heated with a resulting lowering of the resistance of the path and an increase in the current leakage. The heating of the dielectric may lead to rapid deterioration, particularly if moisture is present, and ultimate breakdown. The length of time of the application of the voltage has a definite bearing upon the breakdown voltage. Most dielectrics will withstand for a very brief period a much higher voltage than they can when the voltage is applied for a longer period.

These effects have dictated two tests for capacitors: a high flash-test voltage of very brief duration, and the application of a much lower voltage for a longer period.

The dielectric strength of a material is usually found to be lower for r-f voltages than for a-f or d-c voltages. The rupturing voltage at radio frequencies depends on the rapidity with which the voltage is raised and is not nearly so definite a phenomenon as l-f puncture voltage. Dielectric strength of solid insulators is difficult to measure because of the complexity of the experimental effects. As the r-f currents flow in the material, heating, corona, flashover, and possible deterioration, blistering, or charring may result with consequent changing of voltage and current as the time of application elapses.

If high r-f voltages are applied to an air capacitor, a corona discharge may be set up which appears as a visible glow around high-potential metal parts, points, and sharp edges and is usually distinctly audible. These corona effects represent a power loss. Hence the construction of air capacitors for high voltages requires the rounding of all edges and corners and the avoiding of sharp points which encourage the formation of corona and flashover.

11. Dielectric Absorption. When a capacitor is connected to a d-c source of emf, the instantaneous charge is followed by the flow of a small and steadily decreasing current into the capacitor. The additional charge is absorbed by the dielectric. Similarly the instantaneous discharge of a capacitor is followed by a continuously decreasing current. The capacitor does not become fully charged immediately, nor does it completely discharge immediately when its terminals are shorted, but several discharges may be secured when the capacitor possesses dielectric absorption. The

maximum charge in a capacitor cyclically charged and discharged varies with the frequency of charge.

If a capacitor evidencing dielectric absorption is used at radio frequencies, a power loss occurs which appears as heat. The existence of power loss indicates a component of emf in phase with the current as though a resistance were in series with the capacitor as shown in Fig. 2. The effect of dielectric absorption can be measured along with other losses in the capacitor, although dielectric absorption represents the chief power loss in solid dielectrics.

12. Calculation of Capacitance. Formulas are available for use in calculating the capacitance for a large number of geometrical shapes of conducting surfaces such as spheres and cylinders, either separated or concentric, and flat surfaces of various shapes. The usual types of capacitor calculations are concerned with two or more flat conductors.

When two conducting plates are parallel, close together, and of large area, the capacitance is given by

$$C = 0.0885 \times \frac{KS}{t}$$

where C = capacitance, $\mu\mu\text{f}$

K = dielectric constant (which is 1 for air)

S = area of one plate, sq cm

t = distance between plates, cm

When more than two plates are used, the formula becomes

$$C = 0.0885 \times \frac{KS(N - 1)}{t}$$

where N = number of plates

The actual capacitance of a parallel plate capacitor is slightly larger than the value as calculated from the above formula, because of the fringing of the electric lines of force beyond the space between the plates. A correction¹ can be made for this fringing by slightly increasing the dimensions of the plates. A narrow strip of width w can be added to the actual plate dimensions. In the case of circular plates $w = 0.4413t$, and for plates with straight edges $w = 0.110t$, where t is the distance between the plates in centimeters.

13. Combinations of Capacitors. Combinations of two or more capacitors in a circuit are often arranged in either series or parallel. Capacitors connected in parallel give a total capacitance equal to the sum of the capacitances of the individual capacitors. Capacitors connected in series give a resulting capacitance which may be calculated from the following:

$$C = \frac{1}{\frac{1}{C_1} + \frac{1}{C_2} + \frac{1}{C_3} + \dots}$$

This formula gives the following expression in the case of two capacitors in series

$$C = \frac{C_1 \times C_2}{C_1 + C_2}$$

The various elements such as tubes, sockets, mountings, and wiring in radio apparatus contain many small capacitances by virtue of the difference of potential existing between the numerous conductors insulated from one another. These small capacitances are known as *stray* capacitances. Although they are unimportant in some kinds of work, in other types of work, such as in amplifier design, they must be taken into

¹ Coursey, Philip R., "Electrical Condensers," Sir Isaac Pitman & Sons, Ltd., London.

account. In the case of resistance-coupled amplifiers, for example, these capacitances reduce the amplification at the higher audio frequencies and make a flat characteristic with high over-all gain impossible.

The effect of stray capacitances is eliminated in the case of units used as capacitance standards by shielding the insulated plates and grounding the shield. In this manner a definite capacitance is always assured for a given scale setting.

14. Effect of Frequency on Capacitance. In the best capacitors the effect of frequency is nil in the lower radio range. A criterion of a suitable capacitor for use as a standard is that its capacitance shall be the same for two different sets of charging and discharging conditions. A variable air capacitor, such as the National Bureau of Standards type described on p. 120 of the Bureau's *Circ. 74*, gives the same capacitance at 100 and at 1,000 charges and discharges per second. A capacitor having considerable solid dielectric in its make-up will show a difference in capacitance with frequency. The quantity of electricity that flows into a capacitor during a finite charging period is greater than would flow in during an infinitely short charging period. Consequently the measured or apparent capacitance with a.c. of any finite frequency is greater than the capacitance on infinite frequency, the latter being called the *geometric* capacitance. The capacitance decreases as the frequency increases.

The length of the internal leads should be kept as short and direct as possible to minimize the inductance of the leads that acts to give an apparent change of capacitance with frequency. The amount of this change can be calculated from $C_a = C(1 + \omega^2 CL \times 10^{-12})$, where C_a is the apparent or measured capacitance, C is in microfarads, and L in microhenrys.

15. Capacitor Types. There are many ways in which capacitors might be classified: by their construction, size, voltage rating, use, dielectric, or fixed or variable capacitance. The capacitors used in various radio applications are found in innumerable sizes, shapes, and uses. The two simplest divisions into which they may be classified have to do with their capacitance, i.e., whether it is fixed or variable.

16. Types of Fixed Capacitors. Fixed capacitors are available in all ranges from a few micromicrofarads to several microfarads, for any voltage rating up to 45,000 volts or higher, and in innumerable shapes and dimensions, all depending upon the use for which the unit is intended.

Paper formerly was used as the dielectric for capacitors for use on lower voltages, while mica was used in capacitors for higher voltages. As the art of capacitor manufacture progressed, an oil-impregnated paper dielectric was used in capacitors for the higher voltages, the whole capacitor being mounted within an oil-filled container.

For paper dielectric 100 per cent pure linen paper is used, which must meet severe requirements as to thickness, porosity, uniformity, width, freedom from conducting particles, alkalinity, and acidity. Two or more layers of paper are used between the metal-foil plates, depending upon the voltage for which the capacitor is designed. Paper capacitors are available in hermetically sealed plug-in types to fit standard octal-type radio-tube sockets, both in wax-impregnated and oil-impregnated types for d-c working voltages up to 600.

Paper capacitors are formed by winding two metal-foil electrodes or ribbons in conjunction with the paper ribbons. There are two types of winding: *inductive* and *noninductive*. The latter type is recommended for r-f and for the higher a-f work. The inductive type is satisfactory for l-f work.

In winding the inductive type, the foil used is narrower than the paper, and the contact is made with the foils by tinned copper strips inserted in the winding. The noninductive type of winding is made with the foils about the same width as the paper. The foil is staggered so that the plates project over the ends of the paper. The terminals are soldered to the extending foil at the opposite ends and thus make contact with every turn of the foil. The latter type of construction makes for minimum plate resistance and minimum power loss.

A new type of paper dielectric capacitor has recently appeared, known as a "metal-

ized" paper capacitor. In this type a thin aluminum film from 25 to 100 μm thick is deposited on the paper. This construction reduces the physical size of the capacitor to about one-third that for conventional designs. Plastic film has also been used in some types of capacitors.

Fixed high-voltage vacuum capacitors are available for transmitters and power amplifiers. They consist of the two cylindrical metal elements forming the capacitor mounted within an exhausted glass tube, with metallic terminals on either end to fit into a clip of the cartridge fuse type. These capacitors are made in a number of sizes and ratings from 1 to 5 μmf , 17,000 peak volts, 10 amps, up to 200 μmf , 35,000 peak volts. The physical size of the capacitor increases as the capacitance and peak voltage are increased.

Mica has been used very extensively as dielectric for use at radio frequencies. India mica has been used almost exclusively as it has been generally considered as of superior quality for radio use.

Selected mica is split into sheets of definite thickness, gaged, and tested for punctures or other defects. A capacitor is built up of alternating mica and metal-foil sheets, the sets of plates of opposite polarity being brought out at opposite ends where they are soldered together, forming the two terminals. The whole stack of plates is rigidly clamped together in such a way as to grip the plates firmly in the center and expel all dielectric other than mica.

If a capacitor is to be used with higher voltages, the practice is to construct the capacitor with two or more sections in series, rather than to increase the thickness of the mica. The former method is more flexible than the latter, permitting the construction of capacitors for 45,000 volts or higher.

It is customary to mount the large high-voltage capacitors in steel tanks which are filled with a high flash-point insulating oil which serves to prevent access of dirt and moisture, prevents flashover along the capacitor sections, insulates the capacitor from the tank, and conducts heat away from the capacitor elements.

Table 2 gives the capacitor color code for the 1938 RMA standard, the American war standard, and joint Army-Navy specification. This table and the following material relating to RMA, AWS, and JAN specifications and ceramic capacitors is taken from "Reference Data for Radio Engineers," Federal Telephone and Radio Corporation.

Table 2. Capacitor Color Code for 1938 RMA Standard, American War Standard, and Joint Army-Navy Specification

Color	Significant figure	Capacitance in μmf , decimal multiplier	Tolerance, per cent		Voltage rating, RMA, 1938 standard	Characteristic AWS and JAN mica capacitors
			RMA, 1938 standard	AWS and JAN*		
Black.....	0	1	$\pm 20\text{M}$	A
Brown.....	1	10	1	100	B
Red.....	2	100	2	$\pm 2\text{G}$	200	C
Orange.....	3	1,000	3	300	D
Yellow.....	4	10,000	4	400	E
Green.....	5	100,000	5	500	F
Blue.....	6	1,000,000	6	600	G
Violet.....	7	10,000,000	7	700	
Gray.....	8	100,000,000	8	800	
White.....	9	1,000,000,000	9	900	
Gold.....	..	0.1	± 5	$\pm 5\text{J}$	1,000	
Silver.....	..	0.01	± 10	$\pm 10\text{K}$	2,000	
No color.....	± 20	500	

* Letter used to indicate tolerance in type designations.

AWS and JAN mica capacitors have a characteristic as shown in Table 3.

Table 3

Characteristic	Q	Temperature coefficient, ppm/°C	Max capacitance drift, per cent	Verification of characteristics by production test
A	*	Not specified	Not specified	Not required
B	†	Not specified	Not specified	Not required
C	†	-200 to +200	0.5	Not required
D	†	-100 to +100	0.2	Not required
E	†	0 to +100	0.05	Not required
F	†	0 to + 50	0.025	Required
G	†	0 to - 50	0.025	Required

* Q must be greater than 1/2 of minimum allowable Q for other characteristics (JAN).

† Minimum acceptable Q at 1 Mc is defined by a curve; value varies with capacitance.

Type designations of AWS or JAN fixed mica-dielectric capacitors form a comprehensive numbering system used to identify the component. The capacitor-type designation is given the following form:



Component designation. Fixed mica-dielectric capacitors are identified by the symbol CM.

Case designation. The case designation is a two-digit symbol that identifies a particular case size and shape.

Characteristic. The characteristic is indicated by a single letter in accordance with Table 3.

Capacitance value. The nominal capacitance value in micromicrofarads is indicated by a three-digit number. The first two digits are the first two digits of the capacitance value in micromicrofarads. The final digit signifies the number of zeros that follow the first two digits. If more than two significant figures are required, additional digits may be used, the last digit always indicating the number of zeros.

Capacitance tolerance. The symmetrical capacitance tolerance in per cent is designated by a letter as shown in Table 2.

AWS and JAN Fixed Capacitors. The six-dot color code is shown in Fig. 6

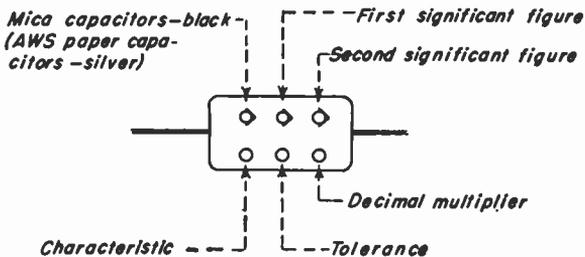


FIG. 6. Six-dot color code for capacitor.

RMA Fixed Capacitors. The 1938 RMA standard covers a simple three-dot color code (Fig. 7) showing directly only the capacitance, and a more comprehensive six-dot color code (Fig. 8) showing three significant figures and tolerance of the capacitance value, and a voltage rating. Capacitance values are expressed in micromicrofarads up to 10,000 $\mu\mu\text{f}$.

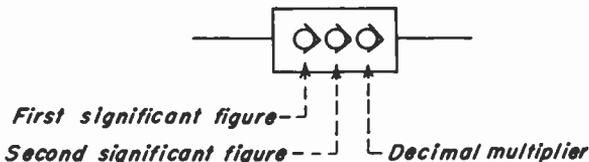


FIG. 7. RMA three-dot 500-volt 20 per cent tolerance code.

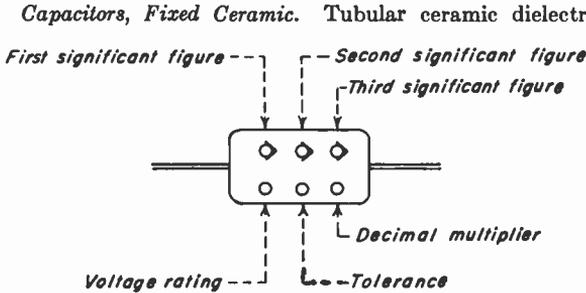


FIG. 8. RMA six-dot capacitor code.

temperature compensation of tuned circuits and have many other applications as well. If the capacitance, tolerance, and temperature coefficient are not printed on the capacitor body, the following color code of Table 4 and Fig. 9 will be used. The change in capacitance per unit capacitance per degree centigrade is the temperature coefficient, usually stated in parts per million per degree Centigrade (ppm/°C).

Table 4

Color	Significant figure	Capacitance, μmf , multiplier	Capacitance tolerance		Temperature coefficient, ppm/°C
			In per cent, $c \leq 10 \mu\text{mf}$	In μmf , $c \geq 10 \mu\text{mf}$	
Black.....	0	1	± 20	2.0	0
Brown.....	1	10	± 1	- 30
Red.....	2	100	± 2	- 80
Orange.....	3	1,000	-150
Yellow.....	4	-220
Green.....	5	± 5	0.5	-330
Blue.....	6	-470
Violet.....	7	-750
Gray.....	8	0.01	0.25	+ 30
White.....	9	0.1	± 10	1.0	-330 ± 500

During the past few years attention has been given by the manufacturers to the development of small capacitors of great stability, whose changes of capacitance with temperature are a definite amount, positive or negative, as desired. Push-button tuned receiving sets required the use of small capacitors that would maintain their capacitance as the receiver warmed up or would change their capacitance to compensate for changes in the coils. A type now available with positive, zero, or negative temperature coefficient employs a small ceramic tube as the dielectric, with silver plating inside and out, followed by copper plating and solder forming the two electrodes, to which wire leads are soldered. Wax impregnation and moistureproof lacquer complete the capacitor, which is said to be unaffected by changes in temperature and humidity. Units of this type have a d-c working voltage of 500 and can be obtained in sizes from 5 to 1,000 μmf .

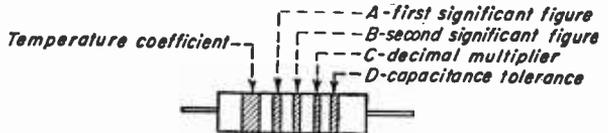


FIG. 9. RMA and AWS capacitor code.

Another type of low-temperature coefficient capacitor uses silver plating on mica and is mounted in either a ceramic or a low-loss bakelite case. These are wax-impregnated and sealed and have small positive temperature coefficients.

Development of ceramics having high dielectric constants has been in progress for

a number of years. Barium-strontium titanates have been found to possess dielectric constants from 10,000 to 18,000. Barium-calcium titanates have lower dielectric constants, and barium-magnesium titanates still lower values. The making of these titanates requires a special process at high temperatures such as 1250 to 1400°C. Capacitors formed of tiny disks of these materials have been used in the construction of "matchbox" receivers and "lipstick" transmitters where all components must be very small.

17. **Electrolytic Capacitors.**¹ Another type of fixed capacitor with important applications is known as the *electrolytic capacitor*. Its advantages are low cost and high capacitance as compared with other types of fixed capacitors. A unit of 8 μ f, 500-volt d-c rating may be manufactured in a tubular assembly $\frac{1}{8}$ in. in diameter by $1\frac{1}{4}$ in. long.

The electrolytic capacitor consists of three essential components: the anode, the dielectric film, and the electrolyte. The anode is always made of aluminum of high purity and forms one plate on the capacitor. The dielectric film is formed electrochemically on the anode and is very thin. The electrolyte may be either a liquid or a pastelike substance. It is the second plate of the capacitor, insulated from the anode plate by virtue of the dielectric film formed on the latter.

Electrolytic capacitors may be divided into two general classes: (1) *dry* electrolytic capacitors in which a pastelike form of electrolyte is used, (2) *wet* electrolytic capacitors in which a liquid or waterlike electrolyte is used.

The electrolyte in the case of dry electrolytic capacitors is absorbed in porous paper and held in position adjacent to the anode foil by this paper. In addition another aluminum foil, generally called the *cathode foil*, is incorporated for the purpose of making electrical contact to the electrolyte-saturated paper.

In the wet type the electrolyte is a dilute water solution. The anode member with adhering dielectric film is suspended in a can, generally of aluminum. The can is then filled with the electrolyte. The can acts as the electrical connection to the electrolyte similar to the cathode foil in the case of the dry electrolytic capacitor.

For a given area of anode surface the capacitance in microfarads is inversely proportional to the thickness of the dielectric film. The film thickness is proportional to the voltage during the electroformation of the film. Therefore, units with very low voltage ratings may be made with very high capacitances. The ordinary ranges are 500 to 6,000 μ f in capacitance for voltage ratings of 6 to 60 volts d.c. and 2 to 100 μ f with voltage ratings of 100 to 150 volts d.c.

18. **Electrolytic Capacitor Characteristics.** The d-c voltage which an electrolytic capacitor can withstand is governed by the voltage at which the original film is applied. It is necessary that the anode always be connected to the positive side of the voltage source. An electrolytic capacitor connected in this manner will operate satisfactorily as long as the applied voltage is of correct polarity and does not exceed rated voltage for more than a few seconds at a time. A reversal of potential will cause the unit to draw

considerable current even at low voltages. A d-c voltage in excess of rated causes the unit to draw an appreciable leakage current.

Dry electrolytic capacitors have a definite breakdown voltage at which permanent failure occurs. Momentary surges less than this breakdown voltage but higher than operating voltage will ordinarily do no damage.

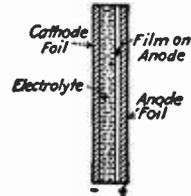


FIG. 10. Electrolytic capacitor construction.

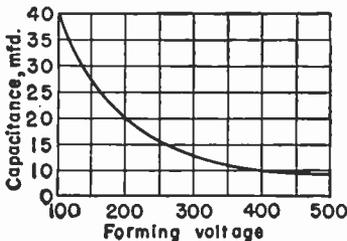


FIG. 11. Electrolytic capacitor characteristic.

¹ Data supplied by E. Sherick, Cornell-Dubilier Electric Corp.

If the anode area is such as to give 8 μf when the working voltage is 500 volts d.c., then the same area at lower working voltages will yield a capacitance as indicated on the curve of Fig. 11.

Electrolytic capacitors have a power factor that is considerably higher than other types of fixed capacitors. This is due in part to the fact that one of the conducting plates is the electrolyte which has considerably higher resistance than the conventional metallic plates of the other types. In effect this places a resistance in series with the capacitor and hence causes a high power factor of the entire unit. Dry electrolytic capacitors have a power factor of about 6 per cent at 60 to 120 cycles. Power factors increase with frequency, and for this reason the use of electrolytic capacitors is generally confined to the l-f application.

19. Etched-foil Types. Within recent years, methods of treating the smooth foil surface in such a way as to make it extremely rough have been applied to electrolytic capacitors. The anode then has an increased total area over and above the original smooth surface. The dielectric film follows the contours of the foil, and the result is

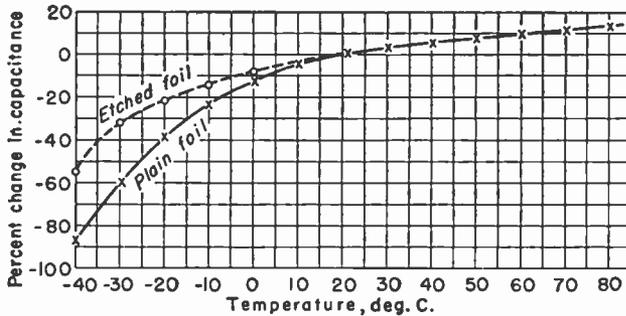


Fig. 12. Comparison of etched- and plain-foil temperature characteristics.

a great increase in capacitance with no increase in volume for any given working voltage.

Several methods have been evolved for formation of a roughened surface for the anode foil. Chief among these are as follows: (1) etching, whereby the smooth foil surface is attacked either chemically or electrochemically and (2) aluminum spraying, whereby the aluminum is sprayed in molten state on a suitable carrier medium.

20. Characteristics under Adverse Conditions. Electrolytic capacitors operate best under normal condition of temperature. The limitations of the electrolyte and the film properties are the governing factors in the operation of this type of capacitor.

Extremely high temperatures cause the electrolyte to dry out and increase in resistivity with consequent increase in power factor when normal temperature is again reached. Furthermore the increase in d-c leakage current with temperature must be considered since there is danger of the start of a destructive cycle due to the generation of internal heat because of the increased d-c leakage.

Temperatures up to 140°F are considered normal although temperatures up to 185°F are not dangerous if the capacitor is rated at 50 to 100 volts higher than the actual operating voltage.

Low temperature causes a decrease in capacitance and an increase in power factor. These changes are temporary and are restored to normal when normal temperatures are again reached.

Where high operating temperatures are to be experienced, the construction should be hermetically sealed in metal cans. This construction limits the loss of electrolyte to a minimum, and longer life is to be expected.

21. Applications. The nature of electrolytic capacitors makes them particularly suitable for filter circuits in power supplies where a relatively high capacitance is

required together with the ability to withstand a d-c potential and small superposed a-c ripple. Second only in importance is the use as a-f by-pass capacitors across screen grids and cathode bias resistors. The use of a-c electrolytics wherein the cathode foil is replaced with a second anode is important in capacitor motor service. These latter units are divided into two classes: those for intermittent duty and those for continuous duty. The former are rated at from 30 to 500 μf at 100 volts a.c. and the latter at 10 to 50 μf at 25 volts a.c.

The intermittent-duty type functions only during the starting of a capacitor motor (capacitor start-induction run). The continuous-duty type functions in the smaller motors rated at about $\frac{1}{100}$ hp and is continuously on voltage during the operation of the motor.

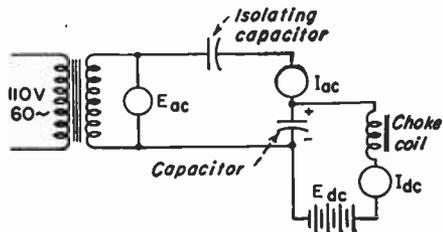


FIG. 13. Production testing circuit for electrolytic capacitors.

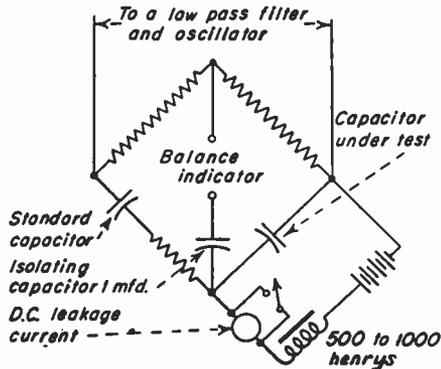


FIG. 14. Circuit for measuring electrolytic capacitors.

23. Types of Variable Capacitors. The most common type of variable capacitor consists of a series of parallel metal plates fastened to a shaft capable of rotation so that the moving plates intermesh with a set of fixed plates. Air is the main dielectric in such capacitors, although some solid insulating material is required to ensure that the two sets of plates are correctly located with respect to each other. Many ways of insulating the plates from each other have been devised, using one or more pieces of the insulating material in sheet, rod, or bar form. Bakelite, hard rubber, pyrex, porcelain, fused quartz, and Isolantite are some of the materials used for such insulators.

The most common use of a variable capacitor is in association with a coil, the combination forming a circuit resonant to a band of radio frequencies depending upon the coil constants and the capacitance range of the capacitor. For a number of applications it is more convenient to have the capacitance change in a different way than

22. Testing. The circuit of Fig. 13 is generally used to test electrolytics in production. E_{dc} supplies a polarizing voltage so that the voltage across the capacitor will be pulsating d.c. The isolating capacitor prevents short-circuiting the polarizing voltage. If E_{dc} is maintained at a constant value, the a-c milliammeter may be calibrated in terms of the capacitance under test. I_{dc} reads the d-c leakage current.

For the accurate measurement of capacitance and power factor, bridge systems such as those shown in Fig. 14 or 15 should be used. They are essentially standard bridge systems rearranged to permit the application of a polarizing voltage.

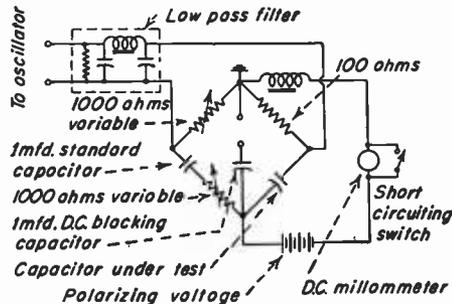


FIG. 15. Capacitance and power-factor measurement.

proportional to the angle of rotation of the plates. This first resulted in the "decrementer" plate and the straight-line wavelength plate. As the use of frequency rather than wavelength became common, the straight-line frequency plate came into use and later the "mid-line" plate. There are other possibilities such as straight-line percentage wavelength and straight-line percentage frequency, the latter being of advantage in frequency measurements. In any of the above shapes or classifications, the movable plates formerly were so shaped as to give the desired frequency or wavelength curve. This resulted in an ill-shaped plate difficult to balance or to hold to a desired setting. In some cases semicircular rotating plates were used with the fixed plates cut away so as to obtain the desired curve. In any of the special forms of plates the plate shape may vary. The minimum and maximum capacitances of the capacitor play a large part in determining the outline of the plate.

Brass or aluminum plates and steel shafts are ordinarily used. If the capacitor is intended for use on high voltages, the spacing between opposite plates must be sufficient to avoid a flash-over or racing between plates. It is customary to round off all sharp edges and corners in such capacitors to avoid flashover.

Capacitors of the air type are often filled with oil, which increases the voltage that they can stand and increases the capacitance from two to five times depending on the dielectric constant of the oil used.

Compressed-air capacitors were formerly used in some radio transmitting stations. The voltage that such a capacitor will stand is increased without changing the capacitance.

Compressed-gas capacitors, utilizing nitrogen under pressure up to 2,000 psi as the dielectric, are now being extensively used in broadcast transmitters. The advantages of low loss and permanent characteristics have long been recognized, but it is only of recent date that any attempt has been made to offer a wide commercial selection of this type of capacitor.

One manufacturer offers three lines with flashover ratings of 15, 20, and 30 kv rms at 1,000 kc, and capacitance ranges up to 1,000, 1,500, and 2,000 μmf , respectively. These are available in fixed, adjustable, or continuously variable types. Special units have been built with flashovers up to 60 kv rms and capacitances up to 20,000 μmf .

Construction varies somewhat with different manufacturers. One offers a completely nonmagnetic assembly using heat-treated aluminum tank and end closures. As a typical example, a variable capacitor having 30 kv rms flashover rating will have a height of 36 in., an over-all diameter of 12 in., and a weight of 90 lb.

Gases other than nitrogen have been used, some of which show considerable promise in increasing flashover voltage and reducing size and weight. These are available in either fixed or variable capacitance types and in sizes from 100 to 2,000 μmf .

Variable high-voltage vacuum capacitors are among the recent developments. These capacitors are similar to the fixed high-voltage vacuum units but have mechanical arrangements to vary the capacitance a few micromicrofarads in the vicinity of 10 or 20 μmf . A larger unit covers a range from 250 to 350 μmf at 10,000 volts peak.

24. Gang Capacitors. The single-dial control radio receiver brought problems to the designer in how to tune two to five circuits accurately using a corresponding number of similar coils and variable capacitors operating on the same shaft. As it is practically impossible conveniently to manufacture two capacitors exactly alike, to say nothing of three or four alike, so that their capacitances shall be exactly the same

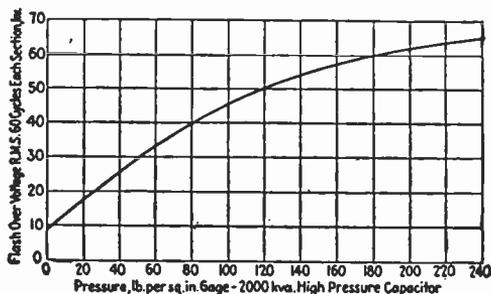


FIG. 16. Flashover voltage (60 cycles) of 2,000-kva capacitor.

throughout the complete rotation of the capacitor plates and accurately tune the capacitors with the same number of similar coils which differ slightly in value, it has been customary to balance or equalize these tuned circuits by the addition of small paralleling capacitors called *trimmer* or *padding* capacitors. Such capacitors can be obtained matched to one-half of 1 per cent. It is possible to obtain two to four capacitors called *gang* capacitors for radio receivers arranged with their shafts in line and operated by one dial, matched to one-half of 1 per cent. The individual capacitors may be separated from one another by metal shields if desired.

Earlier forms of trimmer capacitors used a combination of mica and air as the dielectric. Adjustment in capacitance was accomplished by turning a screw which moved one of the plates nearer the other. A recent design incorporates several concentric cylinders supported on a ceramic member and arranged to interweave so as to change the capacitance from 2.5 to 30 μmf . Another trimmer of recent design employs a metalized glass tube with a metal core within, which may be moved with respect to the metalized coating. Several sizes are available with d-c breakdown voltages from 1,500 to 2,000 volts.

25. Design Equations for Variable Air Capacitors. The capacitance of a capacitor made up of three plates as indicated in Fig. 17 can be obtained by determining the

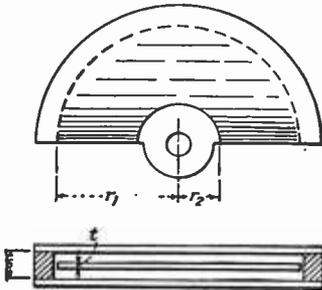


FIG. 17. Dimensions useful in determining capacitance.

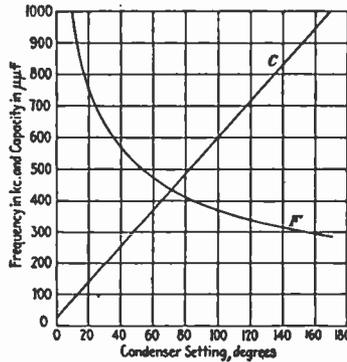


FIG. 18. Semicircular plate characteristic.

area of the overlapping plates, the distance between the adjacent plates, and substitution of these values in the general equation given in Art. 12. The area of the shaded portion of Fig. 17 is $\frac{1}{2}\pi(r_1^2 - r_2^2)$. The distance between the plates is $\frac{1}{2}(s - t)$. Substituting these values in the general equation, the capacitance is given by

$$C = \frac{0.0885 \frac{1}{2}\pi(r_1^2 - r_2^2) \times (3 - 1)}{\frac{1}{2}(s - t)}$$

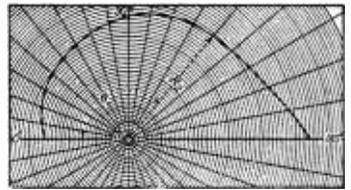
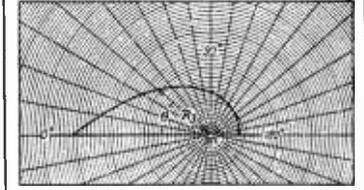
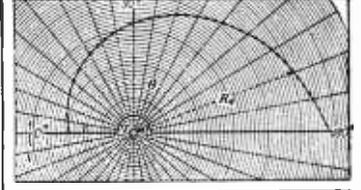
The maximum capacitance of a unit with N plates can be obtained by using a similar equation which may be written

$$C = \frac{0.278(r_1^2 - r_2^2)(N - 1)}{(s - t)}$$

In the above equations C is in micromicrofarads and the dimensions r_1 , r_2 , s , and t in centimeters. These equations neglect the capacitance through the solid insulation which is used in the capacitor and the fringing effect, the correction for which is in Art. 12. Many capacitors are made to have as small a minimum capacitance as possible, giving a large ratio of maximum to minimum capacitance, but this is of doubtful advantage, as slight changes of capacitance due to warping of plates or wear in bearings

will cause a relatively large error at the lower end of the scale but practically no noticeable effect at the maximum capacitance end of the scale.

A semicircular plate capacitor gives a capacitance calibration curve similar to C shown in Fig. 18. With the exception of the portions near the ends of the curve, it is

Straight-line wave length	Straight-line frequency	Straight-line percentage wave length or frequency
		
$C_1 = (a_1\theta + b_1)^2$ $A_1 = k[(a_1\theta + b_1)^2 - \text{resid. cap.}] + K\theta$ $R_1 = [114.6\{2ka_1(a_1\theta + b_1) + K\}]^{\frac{1}{2}}$ <p>Constants:</p> $a_1 = \frac{\sqrt{\text{max. cap.}} - \sqrt{\text{resid. cap.}}}{180}$ $b_1 = \sqrt{\text{resid. cap.}}$	$C_2 = \frac{1}{(a_2\theta + b_2)^2}$ $A_2 = k\left\{\frac{1}{(a_2\theta + b_2)^2} - \text{resid. cap.}\right\} + \frac{K(180 - \theta)}{180}$ $R_2 = [114.6\left\{\frac{2ka_2}{(a_2\theta + b_2)^2} + K\right\}]^{\frac{1}{2}}$ <p>Constants:</p> $a_2 = \frac{1}{180}\left\{\frac{1}{\sqrt{\text{resid. cap.}}} - b_2\right\}$ $b_2 = \frac{1}{\sqrt{\text{max. cap.}}}$	$C_3 = ae^{b_3\theta}$ $A_3 = k[ae^{b_3\theta} - \text{resid. cap.}] + K\theta$ $R_3 = [114.6\{ka_3be^{a_3\theta} + K\}]^{\frac{1}{2}}$ <p>Constants:</p> $a_3 = \text{resid. cap.}$ $b_3 = \frac{\log(\text{max. cap.}) - \log(\text{resid. cap.})}{78.174}$

$$\text{Common constants } k = \frac{\text{total plate area} - 180 K}{\text{max. cap.} - \text{resid. cap.}}$$

$$K = \frac{r^2}{114.6}$$

practically a straight line. In practice, the lower 10 and upper 5 or 10 deg of a 180-deg scale are not used, so as to avoid the curvature in the calibration curve in these regions. Zero setting does not give zero capacitance.

A curve for such a capacitor used with a coil is shown at F in Fig. 18. The frequency changes very rapidly on the lower part of the scale. A slight capacitance change

would make a large frequency change. Therefore, when using frequency meters having semicircular plate capacitors that constitute the main capacitance of the circuit, the coils should be so designed as to give overlaps without resort to the low-capacitance end of the scale.

As the wavelength λ of a wavemeter circuit is proportional to \sqrt{LC} , if L is assumed to be constant, $\lambda \propto \sqrt{C}$ and \sqrt{C} is proportional to the square root of the setting θ . For a uniform wavelength condenser it is necessary to have C vary as the square of the setting θ , or $C \propto \theta^2$.

Again, it may be desirable that the percentage change in capacitance for a given angle of rotation of the plates be the same for all parts of the scale as in the Kolster decimeter.¹ The polar equation for the boundary curve is

$$r = \sqrt{2C_0 a \epsilon^{a\theta} + r_2^2}$$

where C_0 = capacitance when angle $\theta = 0$

a = constant = percentage change of capacitance per scale division

$\epsilon = 2.71828$

r_2 = radius of cutout portion to clear washers separating variable plates

The equations and tables on p. 180 have been compiled by Griffiths.² The four types of plates given are for equivalent capacitors having a capacitance at zero setting of 36 μmf and a maximum of 500 μmf , with a plate area of 20 sq cm.

The paper mentioned above gives the following data for the radii at different angles for the capacitors mentioned in the table of equations.

θ , deg	Radius, cm		
	R_1	R_2	R_4
0	2.49	8.25	1.93
5	2.56		
10	2.60	6.70	2.02
20	2.76	5.62	2.13
30	2.89	4.80	2.24
40	4.17	2.36
60	3.18	3.32	2.64
80	2.75	2.98
90	3.56		
100	2.37	3.38
120	3.86	2.10	3.85
140	1.90	4.40
150	4.12	4.71
160	1.76	5.04
170	5.40
180	4.38	1.65	5.80

26. Effect of Putting Odd-shaped Plate Capacitors in Series or Parallel. If any of the above capacitors are placed in parallel or in series with another capacitor, the straight-line calibration will be altered. If paralleling capacitors are used, the plate shape would require recalculation, after which the plate would become more nearly semicircular. If a capacitor is added in series, the calculation of the plate shape is more difficult. Griffiths³ gives complete equations for a number of series combina-

¹ Nat. Bur. Standards Circ., 74, p. 117. Nat. Bur. Standards Sci. Paper 235.

² Griffiths, W. H. F., Notes on the Laws of Variable Air Condensers, *Exptl. Wireless and Wireless Engr.*, 3, 3-14, January, 1926.

³ Griffiths, W. H. F., Further Notes on the Laws of Variable Air Condensers, *Exptl. Wireless and Wireless Engr.*, 3, 743-755, December, 1926.

tions, the following table applying to the cases indicated where maximum capacitance of variable capacitor = 500 $\mu\mu\text{f}$, minimum capacitance of variable capacitor = 36 $\mu\mu\text{f}$, series fixed capacitance = 500 $\mu\mu\text{f}$, total plate area = 20 sq cm, r = radius of inactive semicircular area of moving plate = 1.2 cm.

θ , deg	Radius, cm			
	R_1	R_2	R_7	R_8
0	2.74	2.16	9.25	1.82
10	2.80	6.95
20	2.35	5.57	1.96
30	2.92	4.65
40	2.56	2.15
50	3.06
60	2.78	3.32	2.38
70	3.22
90	3.40	2.42	2.85
100	3.37
110	3.66
120	2.02	3.57
130	3.88
140	4.25
150	4.18	1.78	4.74
160	4.85
170	4.52
180	4.73	5.66	1.62	7.16

R_1 , straight-line capacitance with series fixed capacitance.

R_2 , corrected square law of capacitance with series fixed capacitance.

R_7 , inverse square law of capacitance with series fixed capacitance.

R_8 , exponential law of capacitance with series fixed capacitance.

27. Important Considerations in Design. It is not difficult to find a large number of capacitors on the market that will answer the needs of any application in radio receivers. The manufacture of capacitors for such use has been brought to a high state of development, both electrically and mechanically. The design problems here are simpler in that low power and low voltage are to be handled.

When capacitors for radio transmitters are designed, provision must be made for handling high power and high voltage. The use of very high radio frequencies has added to the problem by requiring better insulating materials. Insulators that were satisfactory at low radio frequencies have been found to heat up and be unsuited for frequencies such as 30 to 100 Mc and higher.

The following classification shows how capacitors for transmitting sets could be divided with respect to the voltages to which they are subjected:

Those subjected to steady d-c voltages only.

Those subjected to l-f voltages only.

Those subjected to damped r-f voltages only (obsolete)

Those subjected to steady c-w r-f voltages only.

Those subjected to modulated c-w r-f voltages only.

Those subjected to d-c voltages with superimposed r-f voltage.

Those subjected to l-f voltage and superimposed r-f voltage.

The last four of the above divisions could be further subdivided into those for use on frequencies up to about 3,000 kc, those for use on frequencies from 3,000 to about 25,000 kc, and those for use on frequencies of 30,000 kc and above. The two latter classes require special construction.

In specifying the rating of capacitors for use in radio transmitters, the following data should be given: capacitance, current, frequency, nature of voltage to be applied. A knowledge of the maximum r-f voltage and maximum current permissible is important. A capacitor should never be operated at more than half the breakdown voltage. In the case of r-f voltages this fraction should be much smaller.

28. Standards of Capacitance. Fixed capacitors using the best grade of mica or fixed air condensers may be used as capacitance standards for radio frequencies. A variable air capacitor is essential as a standard in many radio measurements.

High-grade mica capacitors can be employed as standards after calibration as to capacitance and power factor over the range of frequencies at which they are to be used.

An important requirement of a standard capacitor is that the capacitance remain constant, the prerequisite of which is rigidity of construction, which is more difficult to secure in a variable than in a fixed capacitor. There should be no relative motion possible between the movable plates and the pointer. There should be no stops against which the pointer or movable plates may strike and thus destroy the calibration. The manner of insulating the two sets of plates is of great importance, not only in fulfilling the rigidity requirement, but in minimizing the power loss. An insulating material having a low temperature coefficient of expansion should be used, so that the capacitance will not change perceptibly with temperature. As small an amount of

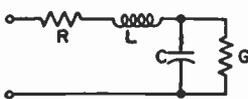


FIG. 19. Equivalent circuit of air capacitor.

solid insulating material as possible should be employed, keeping it well out of the electric field. This field is quite intense near the high-potential post. All insulation should be avoided in the vicinity of that terminal if power factor is to be kept low.

The capacitor should be provided with a metal shield, which may be grounded during measurements if the capacitance is to remain constant. The leads inside the capacitor should be as short and direct as possible. The resistance of leads, plates, and contacts should be kept to the minimum. Flexible connection to the moving plates should not be used in a capacitor used as a standard.

Although it has been customary and is permissible in some measurements to neglect the small resistance and inductance found in variable air capacitors made for precision laboratory work, yet, as the frequency is increased to 5 Mc and higher, such omissions may result in considerable inaccuracy in the results. These small residual impedances, when taken into account, give an equivalent circuit for the variable air capacitor¹ as shown in Fig. 19, where C is the static capacitance of the capacitor, R the resistance loss in the metal parts, L the inductance of the leads and connections of stacks of plates, and G the conductance or losses in the solid dielectric parts.

The resistance R is practically independent of capacitance setting and increases with frequency because of skin effect, being proportional to the square root of the frequency at higher frequencies.

The effect of the inductance L is to cause the apparent capacitance to be greater than the actual capacitance as may be seen from the following equation:

$$C_a = \frac{C}{1 - \omega^2 LC}$$

- where C_a = apparent capacitance, farads
- C = actual capacitance, farads
- L = inductance of capacitor, henrys
- ω = 2π times frequency, cps

The differences between the apparent and the actual capacitances are negligible at medium frequencies but increase rapidly at higher frequencies, where the effect can

¹ Field, R. F., and D. B. Sinclair, A Method for Determining the Residual Inductance and Resistance of a Variable Air Condenser at Radio Frequencies, *Proc. IRE*, **24**, 255-274, February, 1936.

be reduced by feeding the center of the stack of plates rather than the end. This construction is used in variable air capacitors intended for use as standards or for measurements at the higher radio frequencies.

The power factor of a standard variable air capacitor at frequencies up to several megacycles is usually much higher at the low portion of the capacitance scale. As the frequency is raised, the power factor increases very rapidly with increase in scale setting.

29. Methods of Measuring Capacitance. There are two general methods of capacitance measurement: (1) absolute measurements in terms of other electrical or physical units, (2) comparison methods, where a capacitor of unknown capacitance is compared with a known calibrated capacitor. The absolute methods are not carried out at radio frequencies.

Among the simpler comparison methods may be mentioned those operating at an audio frequency or low radio frequency where the measuring device may be direct-reading. Such devices would include capacitor analyzers, bridges, and microfarad meters. Some of these devices are intended for quickly indicating capacitance and may not show more than two significant figures. Precision of measurement is sacrificed for speed in use. Such devices are extremely useful in some cases.

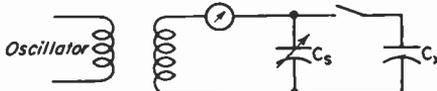


FIG. 21. Simple scheme for measuring capacitance.

There are a number of arrangements that may be used for capacitor calibrations at radio frequencies. These include the simple radio circuits shown in Figs. 20 and 21, r-f bridges, Q meters, and twin-T impedance-measuring circuits. The precision of measurement in any of the various methods cannot exceed that of the standard capacitor.

A simple tuned circuit consisting of a coil and the capacitor under test is arranged with a double-throw switch so that the standard capacitor may be readily substituted. Resonance may be indicated by a sensitive meter coupled to the main coil by a few turns of wire. A crystal detector and 1-ma d-c meter make a very convenient indicating device. Power is supplied electromagnetically by a small vacuum tube oscillator. The measurement circuit is shown in Fig. 20. The shielded side of the capacitor should be grounded. It is essential that the leads connecting the switch points to each capacitor be of the same length in each case, otherwise the circuits will not have the same amount of inductance when one capacitor is substituted for the other. This will result in an error in the calibration. The coupling between the test circuit and the oscillators should be kept quite loose, which will be necessary if a sensitive resonance-indicating instrument is used.

If in the circuit shown in Fig. 20 a fixed inductor is used, the calibration will be made at various frequencies depending upon the capacitance for the different settings. A variable air capacitor of suitable size could be connected across the coil at XX and used to keep the resonance frequency the same for any setting of C_x . If such a circuit is carefully set up, no errors will result if the two circuits connected to C_x and C_s are similar. The frequency at which the measurements are made can be measured with a frequency meter. The frequency or frequency range over which a calibration is made should always be stated.

For rougher calibration work, the circuit shown in Fig. 21 may be used where C_s is tuned both with and without C_x in the circuit. It should be noted that the leads and switch connecting C_x to the circuit will introduce errors in the calibration.

The accuracy of this method can be greatly improved by careful arrangement and shortening of the leads connecting C_x to C_s . The switch should be replaced by a short

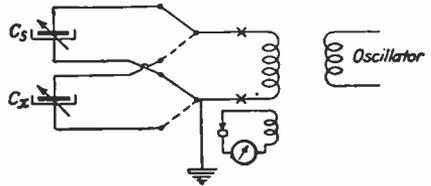


FIG. 20. Measurement of capacitance.

wire, which by the least movement makes or breaks contact with the high-potential terminal of condenser C_x .

A method¹ of precision calibration of variable air capacitors at a single frequency has been described in which the unknown capacitor and the standard capacitor are alternately made a part of the oscillator furnishing the power. The method also offers a very precise means of measuring the change in capacitance with frequency of a mica capacitor.

The manufacturers of equipment for precise measurement of capacitance furnish instructions for use of their instruments. In general, the unknown capacitance is substituted for an equivalent capacitance in the standard capacitor.

The accuracy that can be attained in the measurement of a given capacitor depends upon a number of factors among which are (1) the type of capacitor, *i.e.*, shielded or unshielded, (2) the length of leads employed in connecting the capacitor, (3) the physical dimensions of the capacitor, which also affect the length of leads, (4) the capacitance of the capacitor. Other factors that may be mentioned are accuracy of standard capacitor, ability of test condenser to maintain capacitance value, and temperature effects.

A shielded capacitor is the only type that may be expected to maintain its capacitance value in subsequent use. The capacitance of an unshielded unit may readily be measured, but the value so obtained is affected by its position with respect to grounded panels or other objects, by the terminal that is grounded, lead length, etc., so that unshielded capacitors are not satisfactory where precise values of capacitance are required.

The length of leads, the size, and spacing affect the value of capacitance obtained. The leads should be made as short as possible. Corrections can be made for the leads, or their configuration can be given so that they could be duplicated if desired.

A capacitor of large proportions offers some difficulty because of the long connecting leads required for measurement. Above frequencies of 500 kc, lead corrections should be applied, or procedures used in which the leads are the same length for the standard and unknown capacitors.

Very small and very large capacitors are more difficult to measure accurately than those from 10 to 2,000 or 3,000 μmf . The measurement of very small capacitors is discussed in the next article. Very large capacitors are measured by connecting in series with a smaller capacitor, or by paralleling a number of smaller capacitors. In either case there may be sufficient "stray" capacitances to reduce the accuracy of the measurement.

30. Measurement of Very Small Capacitances. It has always been difficult to obtain agreement between various laboratories in the measurement of capacitances of 5 μmf or smaller. There have been several reasons for this, including differences in the standards of the various laboratories, differences in lead length, size, and spacing, and stray capacitances. Perhaps the greatest difficulty was in the magnitude of the quantity being measured. Consider a capacitance of 1.00 μmf . A 2-in. length of No. 10 wire 0.4 in. above a ground plate would have this capacitance.² A measurement to 1 per cent (0.01 μmf) would, therefore, require very exact positioning of leads, or the use of a jig to ensure exact placement of the condenser under test.

Considerable difficulty was experienced by tube manufacturers early in the recent war, in supplying tubes with interelectrode capacitance within the limits prescribed by the Navy. The difficulty was overcome by the adoption of special test equipment and special measurement techniques.

A new bridge³ has recently been described for measuring interelectrode capacitances at 465 kc. The conductance and capacitance can be measured in terms of standard capacitors having much larger values than the components being measured. Five

¹ Hall, E. L., and W. D. George. Precision Condenser Calibration at Radio Frequencies. *Electronics*, 7, 318-320, 1934.

² Terman, F. E., "Radio Engineers' Handbook," p. 114, McGraw-Hill, 1943.

³ Young, C. H., Measuring Inter-electrode Capacitances. *Tele. Tech.*, 6, 68-70, February, 1947.

decimal divisors are available so that capacitances can be read to 1/100,000 μmf . Bridge balance is sensitive to that amount.

Very small capacitance values or changes in capacitance can be determined by measuring the change in frequency when the unknown capacitance is added to the tuning capacitance of an oscillator. The method described in a previous reference can be adapted by slight rearrangement so as to measure the a-f difference between an auxiliary standard oscillator and the measuring oscillator without the small capacitance and with it in circuit. The small capacitance or change in capacitance is given by the following equation:¹

$$\Delta C_x = C_0 \left[\left(\frac{f_0}{f_1} \right)^2 - 1 \right]$$

where ΔC_x = small capacitance, μmf

C_0 = capacitance of oscillator before ΔC_x was added and includes all capacitances that produce the frequency f_0

f_1 = frequency of oscillator with both C_0 and ΔC_x in parallel

The value of C_0 must be measured. This is done in a similar manner by substituting a known capacitor for ΔC_x and measuring the frequency change with an oscillator.

Capacitance differences can be very precisely measured if a standard is available with a sufficiently finely divided scale. Such standards have taken cylindrical form with a concentric rod, movable longitudinally by means of a fine screw and equipped with a scale to indicate the position of the rod. The change in capacitance can be calculated from the following equation:²

$$C = \frac{0.613}{\log_{10} (b/a)}$$

where C = capacitance, μmf per in.

b = inside diameter of cylinder, in.

a = diameter of inner rod, in.

31. Methods of Measuring Capacitor Resistance and Power Factor and Dielectric Constant of Insulating Materials at Radio Frequencies. Measurements of capacitor resistance and power factor of insulating materials are made in practically the same manner, as the sample of insulating material is prepared so as to form a capacitor. Methods of measuring capacitor resistance³ and power

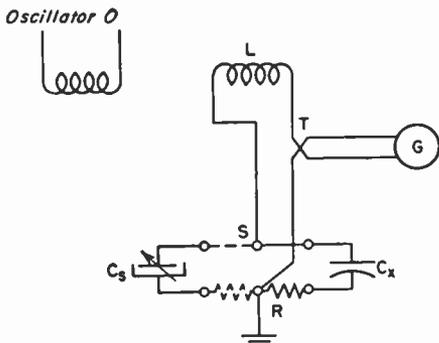


FIG. 22. Circuit for measuring properties of insulators.

factor of insulating materials⁴ have been given in publications of the National Bureau of Standards. The ASTM has several standard methods of testing electrical insulating materials for power factor and dielectric constant.⁵

The circuit shown in Fig. 22 may be used for measurements of resistance, power factor, and dielectric constant. Assuming that the power factor of a sample of insulating material is to be measured, the sample in sheet form is made into a capacitor of capacitance between 100 and 1,000 μmf , as represented by C_x (Fig. 22). The remainder of the circuit consists of the coil L , thermoelement T , and double-pole, double-throw switch S , in which resistors R may be inserted. The galvanometer G gives deflections

¹ Terman, *op. cit.*, p. 924.

² *Ibid.*, p. 922.

³ Radio Instruments and Measurements, *Nat. Bur. Standards Circ. 74*, pp. 190-193.

⁴ Methods of Measurement of Properties of Electrical Insulating Materials, *Nat. Bur. Standards Sci. Paper 471*.

⁵ Tentative Methods of Test for Power Factor and Dielectric Constant of Electrical Insulating Materials, designation D150-46T; Tentative Method of Test for Power Factor and Dielectric Constant of Natural Mica, designation D351-42T; ASTM.

that are proportional to the square of the current flowing in the circuit LTC_1R , as electromagnetically induced from the r-f oscillator O .

The deflections of galvanometer G are noted for several values of inserted resistance R and for the case when R is a link of practically zero resistance. Using the "zero resistance" deflection and the deflection for a known value r of resistance inserted in switch S , the resistance R_T of the total circuit LTC_xR is given by

$$R_T = \frac{r}{\sqrt{\frac{d_0}{d_1}} - 1}$$

The average of the values of R_T calculated for various values of r should be taken as the resistance of the complete circuit. The resistance R_S of the circuit when C_S is substituted for C_X should be obtained in the same manner. The resistance R_X of the capacitor C_x is then given by $R_x = R_T - R_S$. It is essential for this measurement that the two parts of the circuit which are interchanged should be as nearly identical as possible.

After the resistance R_x of the insulating material capacitor is obtained, the power factor or phase difference can be calculated from the equations given above. The dielectric constant K can be calculated from the equation $K = Ct/0.0885S$, where C = capacitance of sample in micromicrofarads, t = thickness of sample in centimeters, and S = area of smaller plate in square centimeters. The capacitance is known, as given by C_s , and the area of one plate and the thickness of the sample can easily be measured.

The method described above operates satisfactorily at frequencies from 100 to 1,500 kc.

The ASTM publication designated D150-46T gives details of procedure for eight methods of measuring dielectric constant and power factor of solid and liquid electrical insulating materials. Four of the methods are for radio frequencies. These are termed (1) the resonant-circuit resistance variation method, with range of 100 to 1,000 kc, (2) the conjugate Schering bridge, with range up to 5,000 kc, (3) the resonant circuit, resonance rise method, with range from 50 to 10,000 kc, and (4) the resonant circuit, susceptance variation method, with range from 100 kc to 100 Mc. Equipment for the latter method is commercially available over part of the lower portion of the range.

Transmission-line and concentric-line methods of impedance measurement can be adapted to power factor and dielectric constant measurements of dielectrics, but there are practical difficulties in the preparation of samples, which limit the usefulness of these methods. A resonant-cavity method¹ has been described operating in the range from 100 to 1,000 Mc which is claimed to have many advantages in use, such as simplicity and rapidity. Several cavities are required, each covering a frequency ratio of 1 to 1.5. Dielectric samples in disk form up to 1½ in. in diameter and from 0.05 to ¼ in. thick are used. From measurements made with the sample in and out of the cavity, the dissipation factor and the dielectric constant are obtained by simple calculation.

¹ Works, C. N., T. W. Dakin, and F. W. Boggs, A Resonant-cavity Method for Measuring Dielectric Properties at Ultrahigh Frequencies, *Trans. AIEE*, 63, 1092-1098, 1944; *Proc. IRE*, 33, 245-254, April, 1945.

CHAPTER 5
 COMBINED CIRCUITS OF *L*, *C*, AND *R*
 BY W. F. LANTERMAN¹

1. Transient and Steady-state Currents. When a voltage is suddenly applied to a circuit, the current assumes a transient state for a brief interval, then gradually settles down to a steady-state condition which it maintains until the voltage is interrupted or changed. Relations for computing transient and steady-state currents and voltages in *LCR* circuits are given in the following paragraphs. The curves shown for current and voltage are typical for each case but are drawn for particular values of the arbitrary constants.

TRANSIENT CURRENTS IN *LCR* CIRCUITS

2. Symbols Used in Transient Expressions. In the transient expressions given in Secs. 3 to 5, the following symbols will be used:

L = inductance, henrys

C = capacitance, farads

R = resistance, ohms

T = time constant, sec; time, sec, for current or voltage to reach $1/\epsilon$ or approximately 36.8 per cent of its initial value if decreasing; or $1 - (1/\epsilon)$, or approximately 63.2 per cent of its final value if increasing

i = instantaneous current, amp, at time *t*

e = instantaneous voltage, volts, at time *t*

e_c = instantaneous capacitor voltage

e_L = instantaneous inductor voltage

t = time, sec, after starting

I = steady-state d.c., amp

E = maximum value of a-c voltage, volts

V = steady-state d-c voltage, volts

Q = capacitor charge, coulombs

Z = a-c impedance, ohms

$$= \sqrt{R^2 + \left(\omega L - \frac{1}{\omega C}\right)^2} \text{ for } LCR \text{ circuit}$$

$$= \sqrt{R^2 + (\omega L)^2} \text{ for } LR \text{ circuit}$$

$$= \sqrt{R^2 + (1/\omega C)^2} \text{ for } RC \text{ circuit}$$

f = frequency of applied a.c., cps

ω = angular velocity of applied a.c., radians per sec = $2\pi f$

f₁ = natural frequency of oscillatory circuit *LCR*, cps

ω_1 = natural angular velocity of oscillatory circuit *LCR*, radians per sec = $2\pi f_1$

W_R = energy, joules, dissipated in *R* during transient state

W_L = energy, joules, stored by or lost by *L* during transient state

W_C = energy, joules, stored by or lost by *C* during transient state

θ = phase angle of a-c voltage at *t* = 0, i.e., when switch is closed

ϕ = phase angle of impedance as defined for each case

ϵ = 2.718 (base of natural logarithms)

α and β = (defined in Sec. 5)

¹ National Broadcasting Co., Inc., Chicago, Ill.

3. RC Circuit Transients.

NOTE: The following formulas for i in RC circuits are not true for extremely small values of t . For very small t , the L of the circuit, no matter how small, limits i , and the relations of Sec. 5 for LCR circuits must be applied. This is especially important for short pulses or high frequencies, where small values of t are involved. The equations for RC circuits neglect the effects of L for small values of t , but the figures show true current forms.

1. D-c Voltage V Suddenly Applied to Deenergized RC.

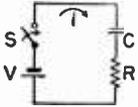


FIG. 1.

$$i = \frac{V}{R} \epsilon^{-(t/RC)}$$

$$e_c^* = V[1 - \epsilon^{-(t/RC)}]$$

$$e_R = V \epsilon^{-(t/RC)}$$

$$T = RC$$

$$W_R = W_C = \frac{1}{2} CV^2 = \frac{1}{2} \frac{Q^2}{C}$$

For fully charged C ($t = \infty$),

$$W_C = W_R = \frac{1}{2} QV$$

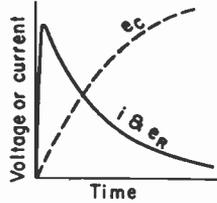


FIG. 2.

2. C Charged to Voltage V and Suddenly Discharged through R.

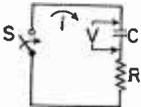


FIG. 3.

$$i = -\frac{V}{R} \epsilon^{-(t/RC)}$$

$$e_c = V \epsilon^{-(t/RC)}$$

$$e_R = -V \epsilon^{-(t/RC)}$$

$$T = RC$$

$$W_C = W_R = \frac{1}{2} CV^2 = \frac{1}{2} \frac{Q^2}{C}$$

For complete discharge of C ($t = \infty$),

$$W_R = \frac{1}{2} QV$$

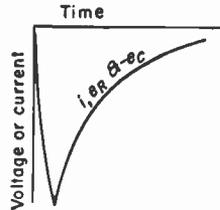


FIG. 4.

3. A-c Voltage e Suddenly Applied to Deenergized RC.

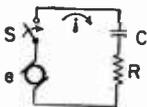


FIG. 5.

Applied voltage $e = E \sin(\omega t + \theta)$

$$i^* = \frac{E}{Z} \sin(\omega t + \theta + \phi)$$

$$- \frac{E}{Z\omega RC} \cos(\theta + \phi) \epsilon^{-(t/RC)}$$

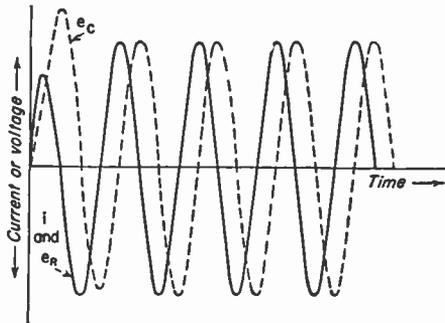


FIG. 6.

$$e_c^* = -\frac{E}{Z\omega C} \cos(\omega t + \theta + \phi) + \frac{E}{Z\omega C} \cos(\theta + \phi) \epsilon^{-(t/RC)}$$

$$e_R^* = \frac{ER}{Z} \sin(\omega t + \theta + \phi) - \frac{E}{Z\omega C} \cos(\theta + \phi) \epsilon^{-(t/RC)}$$

Magnitude of $\phi = \cot^{-1} \omega RC$

NOTE: Negative sense of ϕ has been taken into account in above equations. In substituting, use + value of magnitude for ϕ .

There is no transient if $(\theta + \phi) = 90$ or 270° .

* Underscored terms represent steady-state values; remaining term or terms are the transients.

4. LR Circuit Transients. 1. D-c Voltage V Suddenly Applied to Deenergized LR.

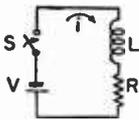


FIG. 7.

$$i^* = \frac{V}{R} [1 - e^{-(Rt/L)}]$$

$$e_L = V e^{-(Rt/L)}$$

$$e_R^* = V [1 - e^{-(Rt/L)}]$$

$$T = \frac{L}{R}$$

$$W_L = \frac{1}{2} LI^2$$

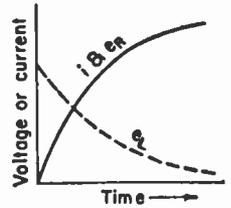


FIG. 8.

For $t = \infty$,

$$W_L = \frac{1}{2} \frac{LV^2}{R^2}$$

2. LR Carrying Steady D-C Suddenly Short-circuited.

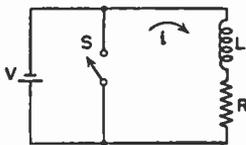


FIG. 9.

Steady current $I = \frac{V}{R}$

$$i = \frac{V}{R} e^{-(Rt/L)} = I e^{-(Rt/L)}$$

$$e_L = -IR e^{-(Rt/L)}$$

$$e_R = IR e^{-(Rt/L)}$$

$$T = \frac{L}{R}$$

$$W_L = \frac{1}{2} LI^2$$

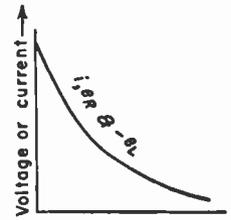


FIG. 10.

For $t = \infty$,

$$W_L = \frac{1}{2} \frac{LV_R^2}{R^2}$$

3. A-c Voltage e Suddenly Applied to Deenergized LR.

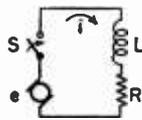


FIG. 11.

Applied voltage $e = E \sin(\omega t + \theta)$

$$i^* = \frac{E}{Z} \sin(\omega t + \theta - \phi)$$

$$- \frac{E}{Z} \sin(\theta - \phi) e^{-(Rt/L)}$$

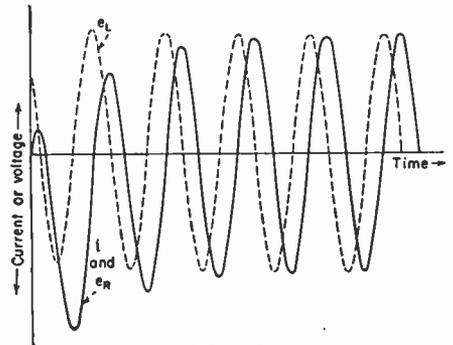


FIG. 12.

$$e_L^* = \frac{E}{Z} \omega L \cos(\omega t + \theta - \phi) + \frac{E}{Z} R \sin(\theta - \phi) e^{-(Rt/L)}$$

$$e_R^* = \frac{ER}{Z} \sin(\omega t + \theta - \phi) - \frac{ER}{Z} \sin(\theta - \phi) e^{-(Rt/L)}$$

$$\text{Magnitude of } \phi = \tan^{-1} \left(\frac{\omega L}{R} \right)$$

There is no transient if $\theta = \phi$ or if $(\theta - \phi) = 180^\circ$.

* Underscored terms represent steady-state values; remaining term or terms are the transients.

5. LCR Circuit Transients. 1. *D-c Voltage V Suddenly Applied to Deenergized LCR.*

General Solutions:

$$i = \frac{V}{2\beta L} \epsilon^{-\alpha t} (\epsilon^{\beta t} - \epsilon^{-\beta t}) = \frac{V}{\beta L} \epsilon^{-\alpha t} \sinh \beta t$$

$$e_c^* = V \left[1 - \epsilon^{-\alpha t} \left(\cosh \beta t + \frac{\alpha}{\beta} \sinh \beta t \right) \right]$$

$$e_L = V \epsilon^{-\alpha t} \left(\cosh \beta t - \frac{\alpha}{\beta} \sinh \beta t \right)$$

$$e_R = 2V \frac{\alpha}{\beta} \epsilon^{-\alpha t} \sinh \beta t$$

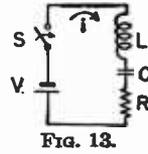


FIG. 13.

where

$$\alpha = \frac{R}{2L}$$

$$\beta = \sqrt{\frac{R^2}{4L^2} - \frac{1}{LC}} = \frac{1}{2L} \sqrt{R^2 - \frac{4L}{C}}$$

There are three special cases, depending upon whether R^2 is greater than, less than or equal to $4L/C$:

Case I. Aperiodic current, when $R^2 > 4L/C$ (β is real). i , e_c , e_L , and e_R are determined directly from the above general solutions by numerical substitution for α and β .

At $t_1 = \left(\frac{1}{2\beta} \log_e \frac{\alpha + \beta}{\alpha - \beta} \right) \text{ sec} = \left(\frac{1}{\beta} \tanh^{-1} \frac{\beta}{\alpha} \right) \text{ sec}$, i and e_R reach maxima, then slowly decay to zero. At the same time, t_1 , e_L becomes zero and thereafter becomes negative, reaching a minimum at $t_2 = \left(\frac{1}{\beta} \log_e \frac{\alpha + \beta}{\alpha - \beta} \right) \text{ sec} = \left(\frac{2}{\beta} \tanh^{-1} \frac{\beta}{\alpha} \right) \text{ sec}$, which is twice the time required for i to reach maximum.

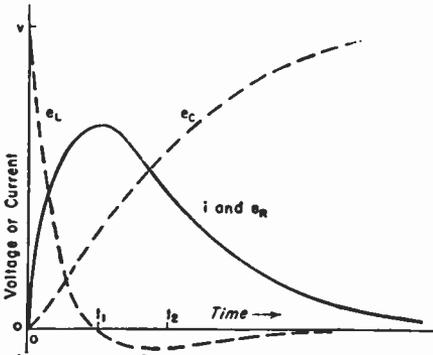


FIG. 14. Case I—aperiodic.

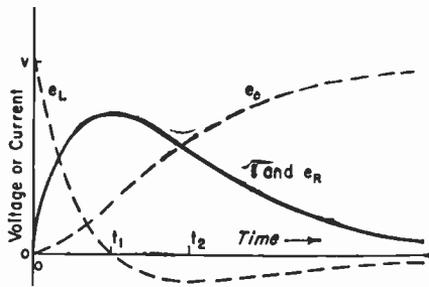


FIG. 15. Case II—critical damping.

Case II. Critical damping, when $R^2 = 4L/C$ ($\beta = 0$).

$$i = \frac{V}{L} t \epsilon^{-\alpha t}$$

$$e_c^* = V [1 - \epsilon^{-\alpha t} (1 + \alpha t)]$$

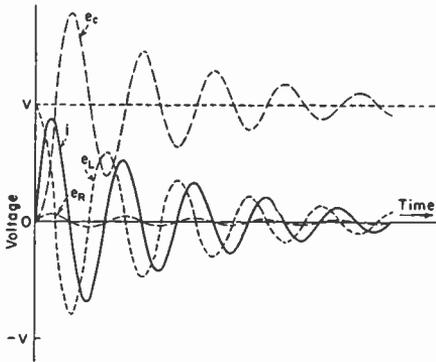
$$e_L = V \epsilon^{-\alpha t} (1 - \alpha t)$$

$$e_R = 2V \alpha t \epsilon^{-\alpha t}$$

At $t_1 = 1/\alpha = 2L/R$ sec, i reaches a maximum of $0.736V/R$ and e_R reaches a maximum of $0.736V$. At the same time, t_1 , e_L becomes zero and thereafter becomes negative, reaching a minimum of $-0.1353V$ at $t_2 = 2/\alpha$ sec, which is twice t_1 .

* Underscored terms represent steady-state values; remaining term or terms are the transients.

Case III. Oscillatory current, when $R^2 < 4L/C$ (β is imaginary).



$$i = \frac{V}{\omega_1 L} \epsilon^{-\alpha t} \sin \omega_1 t$$

where $\omega_1 = \sqrt{\frac{1}{LC} - \frac{R^2}{4L^2}} = 2\pi f_1$

$$e_C^* = V \left[1 - \epsilon^{-\alpha t} \left(\cos \omega_1 t - \frac{\alpha}{\omega_1} \sin \omega_1 t \right) \right]$$

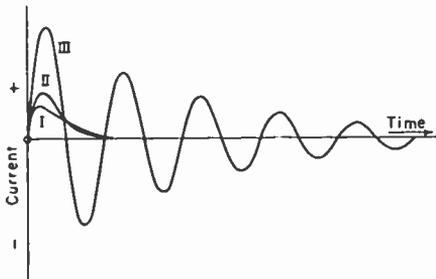
$$e_L = V \epsilon^{-\alpha t} \left(\cos \omega_1 t - \frac{\alpha}{\omega_1} \sin \omega_1 t \right)$$

$$e_R = 2V \epsilon^{-\alpha t} \frac{\alpha}{\omega_1} \sin \omega_1 t$$

NOTE: $\frac{\alpha}{\omega_1} = \sqrt{\frac{R^2 C}{4L - R^2 C}}$

FIG. 16. Case III—oscillatory. At $t_1 = 1/4f_1$ sec, approximately, i reaches a maximum and at $t_2 = 1/2f_1$ sec, e_C reaches a maximum. e_L is maximum at $t = 0$. Beyond their respective maxima all four quantities oscillate at frequency f_1 , with amplitudes slowly decreasing to zero, except e_C , which approaches V .

Superimposed Curves. Curves for i in the three special cases are shown superimposed for direct comparison in Fig. 17.



The conditions for all three cases are identical except that R is varied to alter the value of α .

2. A-c Voltage e Suddenly Applied to Deenergized LCR.

Applied voltage $e = E \sin(\omega t + \theta)$

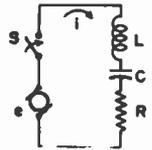


FIG. 17. Superimposed current curves for d.c. in LCR.

FIG. 18.

There are three special cases, depending upon the ratio

$$\alpha = R/2L$$

Case I. Aperiodic current, when $\alpha^2 = R^2/4L^2 > 1/LC$ (β is real).

$$i^* = \frac{E}{Z} \sin(\omega t + \theta - \phi) + \frac{E}{Z} \frac{1}{2\beta L} \left[L(\alpha - \beta) \sin(\theta - \phi) - \frac{1}{\omega C} \cos(\theta - \phi) \right] \epsilon^{-(\alpha - \beta)t} - \frac{E}{Z} \frac{1}{2\beta L} \left[L(\alpha + \beta) \sin(\theta - \phi) - \frac{1}{\omega C} \cos(\theta - \phi) \right] \epsilon^{-(\alpha + \beta)t}$$

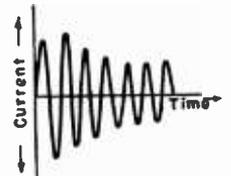


FIG. 19. Case I—aperiodic.

* Underscored terms represent steady-state values; remaining term or terms are transients.

Case II. Critical damping, when $\alpha^2 \equiv R^2/4L^2 = 1/LC$ ($\beta = 0$).

$$i^* = \frac{E}{Z} \sin(\omega t + \theta - \phi) + \frac{E}{Z} \left[(\alpha t - 1) \sin(\theta - \phi) - \frac{1}{L\omega C} \cos(\theta - \phi) \right] e^{-\alpha t}$$

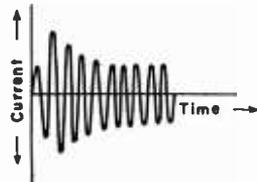
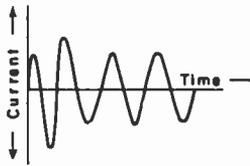


FIG. 20. Case II—critical damping.

Case III. Oscillatory current, when $\alpha^2 \equiv R^2/4L^2 < 1/LC$ (β is imaginary).



$$i^* = \frac{E}{Z} \sin(\omega t + \theta - \phi) - \frac{E}{Z} e^{-\alpha t} \left[\sin(\theta - \phi) \cos \omega t + \frac{1}{\omega L \omega C} \cos(\theta - \phi) \sin \omega t - \frac{R}{2\omega L} \sin(\theta - \phi) \sin \omega t \right]$$

FIG. 21. Case III—oscillatory.

6. Time Constants. Time constant is an arbitrary factor used to gage the exponential rate of transient current or voltage increase in a circuit containing reactance and resistance in series. It denotes only the relative duration of the transient, and no special physical phenomenon occurs when the transient has existed for a length of time equal to the time constant.

For exponentially increasing quantities the time constant is the time in seconds required for the quantity to attain a value equal to $1 - (1/e) = 0.63212$ (or 63.2 per cent, approximately) of its final value. For exponentially decreasing quantities, it is the time in seconds required for the quantity to attain $1/e = 0.36788$ (or 36.8 per cent, approximately) of its initial value.

In series LR circuits, the time constant is $T = L/R$ sec; in series RC circuits, $T = RC$ sec; where R is in ohms, L in henrys, and C in farads (see Secs. 3 and 4).

7. Logarithmic Decrement. The logarithmic decrement is the natural logarithm of the ratio of amplitudes of two successive oscillations (of the same polarity) of a damped wave train. In terms of $L, R, C,$ and $Q,$ the decrement is

$$\delta = \frac{2\pi R}{\sqrt{4(L/C) - R^2}} = \frac{R}{2Lf_r} = \frac{\pi}{Q} \quad (1)$$

where f_r is the resonant frequency of the circuit, and $Q = \omega_r L/R$.

8. Pulsed Circuits. In television synchronizing systems and other timing circuits the input voltage applied to networks of the types shown in Fig. 22 consists of a series of one or more d-c pulses of the forms shown in Fig. 23. Idealized, the pulse voltage is zero up to a given instant and, at each instant thereafter up to time $= P$ sec, it has a succession of finite values determined by some given pulse form. At P sec, the voltage drops instantly to zero and remains equal to zero until the start of the next pulse. The networks and pulses in Figs. 22 and 23 are typical only; any

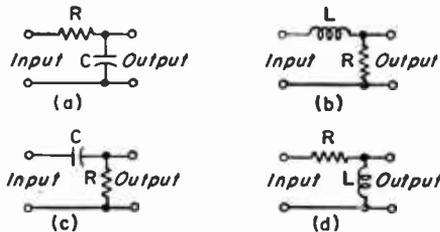


FIG. 22. Typical networks used with pulses.

* Underscored terms represent steady-state values; remaining term or terms are the transients.

one of an infinite number of pulse shapes may be applied to any one of an infinite number of networks.

Solutions for the output voltage or current are obtained by considering the pulse as creating a transient condition starting at $t = 0$, followed by a new transient condition (to restore the input voltage to zero) starting at $t = P$. If the pulse form can be expressed as a function of t , $e = f(t)$, the output voltage or current is obtained by setting up and solving the differential equations in which the sum of the instantaneous voltage drops in the circuit is equated to $f(t)$. At $t = P$ there is a discontinuity where e suddenly decreases to zero, if the pulse shape is not such that e is already zero when $t = P$. This requires solution of a new differential equation in which the voltage drops existing in the circuit at $t = P$ are equated to zero. This equation must take into account the energy stored in the inductances and capacitances at the instant when $t = P$.

FIG. 23. Typical pulse forms.

Unfortunately, each pulse circuit and each pulse shape is a special case, and results cannot be generally formulated except for typical examples.

Example: For a saw-tooth pulse of the form of Fig. 23b applied to the LR network of Fig. 22d the differential equation applying between $t = 0$ and $t = P$ is

$$L \frac{di}{dt} + iR = Vat \tag{2}$$

in which $i = 0$ when $t = 0$, and V and a are defined in Fig. 25. The solution is

$$i = \frac{Vat}{R} - \frac{VaL}{R^2} [1 - e^{-(Rt/L)}] \tag{3}$$

and the output voltage is

$$e_L = L \frac{di}{dt} = \frac{VaL}{R} [1 - e^{-(Rt/L)}] = V_n [1 - e^{-(t/nP)}] \tag{4}$$

where symbols are as defined in Fig. 25.

At $t = P$ the circuit is short-circuited and corresponds to an LR circuit carrying d.c. I and suddenly short-circuited as in Sec. 4b. The voltage equation for $t > P$ is

$$L \frac{di}{d(t - P)} + iR = 0 \tag{5}$$

in which $i = I$ when $t = P$. The solution is

$$i = I e^{-[R(t-P)/L]} \tag{6}$$

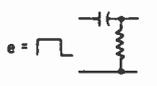
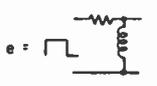
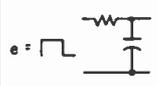
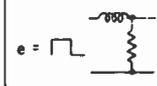
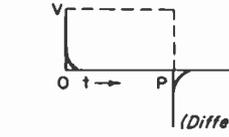
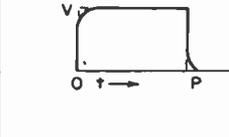
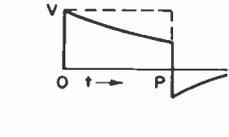
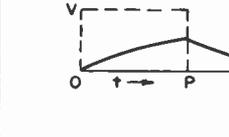
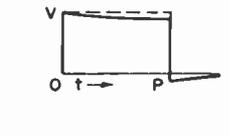
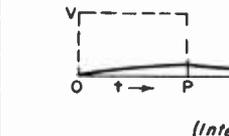
and the output voltage is

$$e_L = L \frac{di}{d(t - P)} = -IR e^{-[R(t-P)/L]} = -V [1 - n(1 - e^{-1/n})] e^{-(t-P)/nP} \tag{7}$$

In any practical case, the voltage drop in the source and the effect of the load upon the pulse output cannot be ignored. That is, the L 's, C 's, and R 's in the source and load must be considered as parts of the network. To preserve the pulse shape through a network, it is customary to employ a low-impedance source and a high-impedance load.

Figures 24 to 27 show output voltages for the four pulse forms of Fig. 23 applied to

the networks of Fig. 22. In this case, it is assumed that the input is short-circuited at $t = P$ and thereafter; also that the load impedance is very large compared to that of the network element from which the output voltage is taken. It is also assumed that the Q 's of L and C are sufficiently large so that their resistances are negligible as compared to the R 's of the network.

Circuit						
Time constant		RC	L/R	RC	L/R	
e_{out}	$t < P$	$V \epsilon^{-\frac{t}{nP}}$			$V(1 - \epsilon^{-\frac{t}{nP}})$	
	$t = P$	$V \epsilon^{-\frac{1}{n}}$			$V(1 - \epsilon^{-\frac{1}{n}})$	
	$t > P$	$-V(1 - \epsilon^{-\frac{1}{n}}) \epsilon^{-\frac{t-P}{nP}}$			$V(1 - \epsilon^{-\frac{1}{n}}) \epsilon^{-\frac{t-P}{nP}}$	
e_{out}	$T \ll P$ $n \rightarrow 0$					
	$T = P$ $n = 1$					
	$T \gg P$ $n \rightarrow \infty$					

Symbols and relations used:
 t = time, sec, from start of pulse
 T = time constant, sec
 P = pulse duration, sec
 V = maximum pulse voltage (input)
 $n = T/P$ (or $T = nP$)

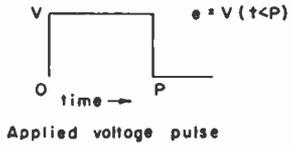


FIG. 24. Equations and pulse forms for output voltage of pulsed networks with rectangular pulses of duration P sec applied.

The exponential pulse forms are chiefly of interest when the origin of the input pulse to a particular network is a preceding network whose output pulse has the exponential form. For example, the output pulse is exponential when a rectangular pulse is applied to any of the four networks of Fig. 22.

9. Differentiating and Integrating Networks. Circuits of the types of Fig. 22c and d, in which the time constant is made very small as compared to the pulse length,

produce output voltage pulses that are approximately the derivatives of the input voltage pulses, as shown in Fig. 28. From this fact, they derive the name *differentiating networks*.

Similarly, circuits of the types of Fig. 22a and b, in which the time constant is made

Circuit					
Time constant		RC	L/R	RC	L/R
e_{out}	$t < P$	$Vn(1 - e^{-\frac{t}{nP}})$		$V[\frac{t}{P} - n(1 - e^{-\frac{t}{nP}})]$	
	$t = P$	$Vn(1 - e^{-\frac{1}{n}})$		$V[1 - n(1 - e^{-\frac{1}{n}})]$	
	$t > P$	$-V[1 - n(1 - e^{-\frac{1}{n}})]e^{-\frac{t-P}{nP}}$		$V[1 - n(1 - e^{-\frac{1}{n}})]e^{-\frac{t-P}{nP}}$	
e_{out}	$T \ll P$ $n \rightarrow 0$				
	$T = P$ $n = 1$				
	$T \gg P$ $n \rightarrow \infty$				

Symbols and relations used:
 t = time, sec, from start of pulse
 T = time constant, sec
 P = pulse duration, sec
 V = maximum pulse voltage (input)
 $n = T/P$ (or $T = nP$)
 a = arbitrary constant

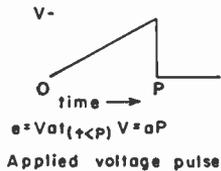


FIG. 25. Equations and pulse forms for output voltage of pulsed networks with saw-tooth pulses of duration P sec applied.

large compared to the pulse duration, produce output voltage pulses that are approximately the integrals of the input voltage pulses, as shown in Fig. 29. This type of network is called an *integrating network*.

10. "Constant-current" RC Circuit. In a series RC circuit as shown in Fig. 30a, when C is discharged through R , the current varies logarithmically with time (see

Fig. 4, Sec. 3b). For pulses of short duration relative to the time constant R_0C an approximation to constant current is obtained by inserting the network $L-R$ as in Fig. 30b, where the time constant L/R is made much larger than the pulse duration time. L and R are determined by the relation $L = R^2C$, and the resulting current is

Circuit					
Time constant		RC	L/R	RC	L/R
e_{out}	$t < P$	$\alpha = 1: V \frac{1}{nP} \epsilon^{-\frac{t}{nP}}$	$\alpha \neq 1: V \alpha \left[\epsilon^{-\frac{\alpha t}{nP}} - \epsilon^{-\frac{t}{nP}} \right] / (1 - \alpha)$	$\alpha = 1: V \left[1 - \left(1 + \frac{t}{nP}\right) \epsilon^{-\frac{t}{nP}} \right]$	$\alpha \neq 1: V \left[1 - \frac{\alpha t}{nP} - \alpha \epsilon^{-\frac{t}{nP}} \right] / (1 - \alpha)$
	$t = P$	$\alpha = 1: V \frac{1}{n} \epsilon^{-\frac{1}{n}}$	$\alpha \neq 1: V \alpha \left[\frac{\epsilon^{-\frac{\alpha}{n}} - \epsilon^{-\frac{1}{n}}}{1 - \alpha} \right]$	$\alpha = 1: V \left[1 - \left(1 + \frac{1}{n}\right) \epsilon^{-\frac{1}{n}} \right]$	$\alpha \neq 1: V \left[1 - \frac{\alpha}{n} - \alpha \epsilon^{-\frac{1}{n}} \right] / (1 - \alpha)$
	$t > P$	$\alpha = 1: -V \left[1 - \left(1 + \frac{1}{n}\right) \epsilon^{-\frac{1}{n}} \right] \epsilon^{-\frac{t-P}{nP}}$	$\alpha \neq 1: -V \alpha \left[1 - \frac{\alpha}{n} - \alpha \epsilon^{-\frac{1}{n}} \right] \epsilon^{-\frac{t-P}{nP}}$	$\alpha = 1: V \left[1 - \left(1 + \frac{1}{n}\right) \epsilon^{-\frac{1}{n}} \right] \epsilon^{-\frac{t-P}{nP}}$	$\alpha \neq 1: V \left[1 - \frac{\alpha}{n} - \alpha \epsilon^{-\frac{1}{n}} \right] \epsilon^{-\frac{t-P}{nP}}$
e_{out} (based on $a=1$)	$T \ll P$ $n \rightarrow 0$				
	$T = P$ $n = 1$				
	$T \gg P$ $n \rightarrow \infty$				

Symbols and relations used:
 t = time, sec, from start of pulse
 T = time constant, sec
 P = pulse duration, sec
 V = maximum pulse voltage (input)
 $n = T/P$ (or $T = np$)
 a = arbitrary constant
 $\alpha = aT$

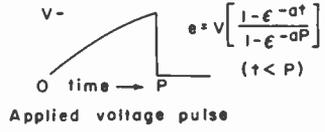


FIG. 26. Equations and pulse forms for output voltage of pulsed networks with exponentially increasing pulses of duration P sec applied.

essentially independent of the load resistance R_0 except in magnitude. The actual variation of the value of i during such a pulse is reduced to about 10 per cent of the variation obtained with the simple R_0C circuit.¹

¹ Fundingsland, O. T., and G. J. Wheeler, *Electronics*, November, 1946, p. 130.

Circuit				
Time constant	RC	L/R	RC	L/R
e_{out}	$t < P$	$\alpha = 1: V(1 - \frac{t}{nP})\epsilon - \frac{t}{nP} \left \alpha \neq 1: V \left[\frac{\epsilon^{-\frac{t}{nP}} - \alpha \epsilon^{-\frac{\alpha t}{nP}}}{1 - \alpha} \right] \right.$	$\alpha = 1: V \frac{t}{nP} \epsilon - \frac{t}{nP} \left \alpha \neq 1: \frac{\alpha t}{nP} \left[\frac{\epsilon^{-\frac{t}{nP}} - \epsilon^{-\frac{\alpha t}{nP}}}{1 - \alpha} \right] \right.$	
	$t = P$	$\alpha = 1: V(1 - \frac{1}{n})\epsilon^{-\frac{1}{n}} \left \alpha \neq 1: V \left[\frac{\epsilon^{-\frac{1}{n}} - \alpha \epsilon^{-\frac{\alpha}{n}}}{1 - \alpha} \right] \right.$	$\alpha = 1: V \frac{1}{n} \epsilon^{-\frac{1}{n}} \left \alpha \neq 1: V \left[\frac{\epsilon^{-\frac{1}{n}} - \epsilon^{-\frac{\alpha}{n}}}{1 - \alpha} \right] \right.$	
	$t > P$	$\alpha = 1: -V \frac{1}{n} \epsilon^{-\frac{t}{nP}} \left \alpha \neq 1: -V \left[\frac{\epsilon^{-\frac{\alpha}{n}} - \epsilon^{-\frac{1}{n}}}{1 - \alpha} \right] \epsilon^{-\frac{t-P}{nP}} \right.$	$\alpha = 1: V \frac{1}{n} \epsilon^{-\frac{t}{nP}} \left \alpha \neq 1: V \left[\frac{\epsilon^{-\frac{\alpha}{n}} - \epsilon^{-\frac{1}{n}}}{1 - \alpha} \right] \epsilon^{-\frac{t-P}{nP}} \right.$	
e_{out} (based on $a=1$)	$T \ll P$ $n \rightarrow 0$			
	$T = P$ $n = 0$			
	$T \gg P$ $n \rightarrow \infty$			

Symbols and relations used:
t = time, sec, from start of pulse
T = time constant, sec
P = pulse duration, sec
V = maximum pulse voltage (input)
n = *T/P* (or *T = nP*)
a = arbitrary constant
 $\alpha = aT$

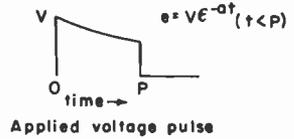


FIG. 27. Equations and pulse forms for output voltage of pulsed networks with exponentially decreasing pulses of duration *P* sec applied.

STEADY-STATE CURRENTS IN LCR CIRCUITS

11. *Q* in LCR Circuits. By definition,

$$Q = \frac{\text{volt-amp stored}}{\text{watts dissipated}} = \frac{X}{R} \tag{8}$$

Q of a coil:

$$Q = \frac{\omega L}{R} = \frac{2\pi f L}{R} \tag{9}$$

Q of a capacitor:

$$Q = \frac{1}{\omega CR} = \frac{1}{2\pi fCR} \tag{10}$$

Q of a circuit:

$$Q = \frac{\omega L}{R_t} \quad \text{or} \quad \frac{1}{\omega CR_t} \tag{11}$$

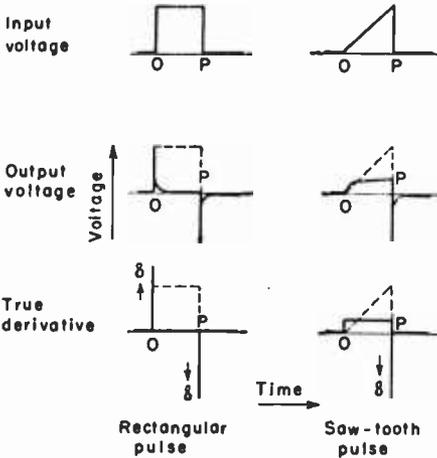


FIG. 28. Pulse forms produced by differentiating networks.

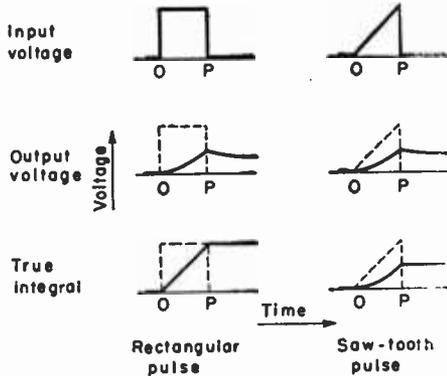


FIG. 29. Pulse forms produced by integrating networks.

where R_t is the total resistance and L or C is the total inductance or capacitance in the circuit. If both L and C are present, $\omega = 2\pi \times$ the resonance frequency.

As a rule, the Q of a coil or capacitor varies less with frequency than does the effective resistance. Thus the assumption that Q is constant is usually more nearly accurate than the assumption that R is constant in expressions for tuned circuits. Expressions containing Q are still approximations, however, and are often loosely used without regard for the fact that Q actually does vary with frequency. Such expressions are most nearly correct for large values of Q (of the order of 100) and when the frequency range is small.

In general, the "sharpness" of resonance of a circuit increases with increased values of Q .

Table 1 gives some representative values of Q that may be expected from coils and capacitors of types commonly used within the frequency ranges shown.

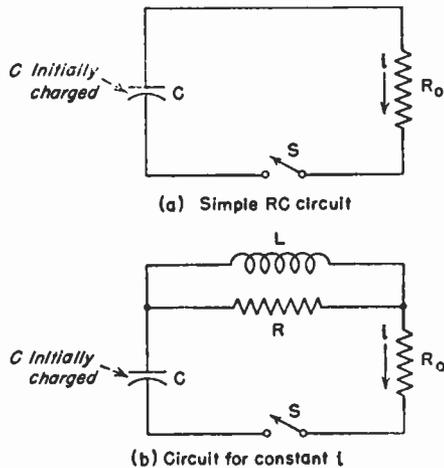


FIG. 30. Modified RC circuit for short pulses of constant current from discharge of C .

Table 1. Representative Values of *Q* for Various Coils and Capacitors

Frequency, cycles	Coils with powdered iron cores	Air-cored coils	Capacitors with paper dielectric	Capacitors with mica dielectric
100	25- 50	3- 10	1,000	
1,000	50- 75	25- 50	500	3,000
10,000	100-150	100-300	100-200	500
100,000	150-200	100-300	50-100	200-300
1,000,000	100-200	100-300	50-200

The following data are quoted from Franks:¹

Item	Frequency, kc	<i>Q</i>	Item	Frequency, kc	<i>Q</i>
100 μf molded bakelite fixed capacitor.....	1,000	40	Broadcast band bank-wound litz solenoid ¾ in. diam in 1½-in. square shield can.....	1,000	110
Typical gang capacitor; bakelite stator insulation.	100	2,000	Broadcast band universal-wound litz coil with iron core in same can.....	1,000	185
	1,000	700	Transmitter coil, 4½ in. diam and 5 in. long, 11 turns of ¼-in. copper tubing.....	5,000	650
Same with ceramic stator insulation.	10,000	200	Transmitter coil 1¾ in. diam and 1¾ in. long, 12 turns of No. 10 wire.....	10,000	400
	100	8,000	Receiver coil, 1 in. diam and 5½ in. long, 5 turns of No. 14....	30,000	270
	1,000	3,000			
	10,000	1,000			
Single-section litz-wound universal coil; 456-kc i.f., in can.	456	80			
Same but with powdered-iron core.....	456	145			

12. A-c Impedance Relations. For applied voltage, $e = E \sin \omega t$, $\omega = 2\pi f$, f = frequency in cycles.

$$Z = \frac{E}{I} = R + jX = |Z| \angle \phi \quad \text{ohms} \tag{12}$$

$$Z = \sqrt{R^2 + X^2} \quad \text{ohms} \tag{13}$$

$$\phi = \tan^{-1} \frac{X}{R} = \cos^{-1} \frac{R}{Z} = \sin^{-1} \frac{X}{Z} \quad \text{radians or deg} \tag{14}$$

Reactance of *C*:

$$X_C = -\frac{1}{\omega C} = -\frac{1}{2\pi f C} \quad \text{ohms} \tag{15}$$

Reactance of *L*:

$$X_L = \omega L = 2\pi f L \quad \text{ohms} \tag{16}$$

Impedance of *LCR* in series:

$$Z = R + j \left(\omega L - \frac{1}{\omega C} \right) \quad \text{ohms} \tag{17}$$

Impedances in series:

$$Z_0 = Z_1 + Z_2 + Z_3 + \dots \quad (\text{see Fig. 31}) \tag{18}$$

¹ Franks, C. J., *Electronics*, April, 1935, p. 126.

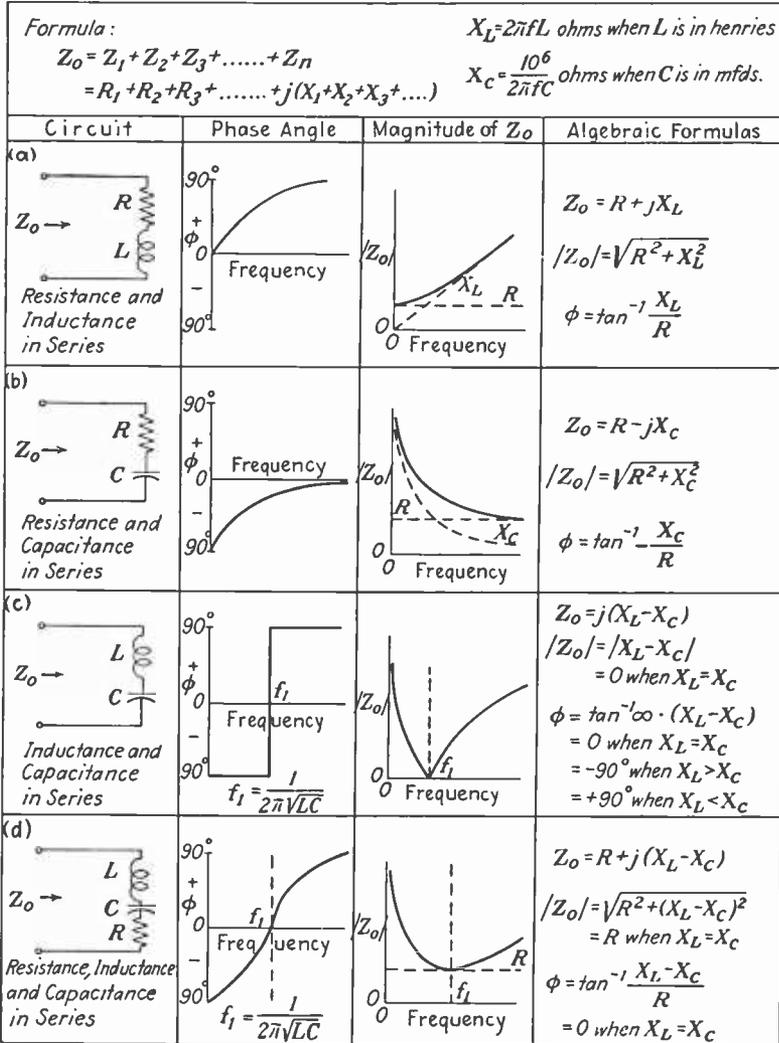


FIG. 31. Equivalent impedances of series combinations of L, C, and R.

Impedances in parallel:

$$Z_0 = \frac{Z_1 Z_2}{Z_1 + Z_2} \quad (\text{see Fig. 32a and b}) \quad (19)$$

13. Loss Due to Inserting Series or Shunt Impedance in Audio Circuits. In audio circuits, attenuation-frequency characteristics are often purposely modified by the insertion of corrective impedances such as equalizers, "tone controls," and scratch filters. The following formulas give the insertion losses in such cases:

1. Shunt Impedance. The loss due to inserting a shunt impedance Z (Fig. 33a and b) is

$$L = 20 \log_{10} \left(1 + \frac{Z_1 Z_2}{Z_s(Z_1 + Z_2)} \right) \text{ db} \tag{20}$$

The shunting impedance can usually be located at a point in the circuit where the impedances Z_1 and Z_2 are matched and where each is substantially a pure resistance through the

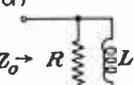
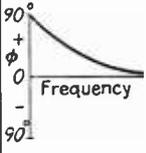
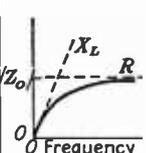
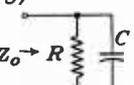
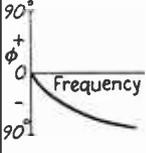
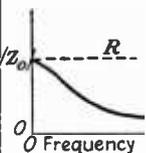
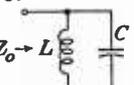
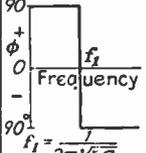
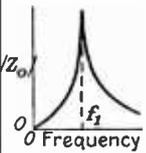
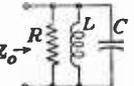
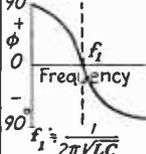
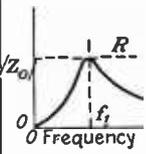
Circuit	Phase Angle	Magnitude of Z_o	Algebraic Formulas
 <p>$Z_o \rightarrow R$ and L in parallel.</p> <p>Inductance and Resistance in Parallel</p>			$Z_o = \frac{RX_L(X_L + jR)}{R^2 + X_L^2}$ $ Z_o = \frac{RX_L}{\sqrt{R^2 + X_L^2}}$ $\phi = \tan^{-1} \frac{R}{X_L}$
 <p>$Z_o \rightarrow R$ and C in parallel.</p> <p>Resistance and Capacitance in Parallel</p>			$Z_o = \frac{RX_C(X_C - jR)}{R^2 + X_C^2}$ $ Z_o = \frac{RX_C}{\sqrt{R^2 + X_C^2}}$ $\phi = \tan^{-1} \frac{R}{X_C}$
 <p>$Z_o \rightarrow L$ and C in parallel.</p> <p>Inductance and Capacitance in Parallel</p>	 <p>$f_1 = \frac{1}{2\pi\sqrt{LC}}$</p>		$Z_o = -j\frac{L}{C(X_L - X_C)}$ $ Z_o = \left \frac{L}{C(X_L - X_C)} \right $ <p>= ∞ when $X_L = X_C$ = 0 when $X_L = X_C$</p> $\phi = \tan^{-1} \infty \cdot \left(\frac{-X_L X_C}{X_L - X_C} \right)$
 <p>$Z_o \rightarrow R, L, C$ in parallel.</p> <p>Resistance, Capacitance and Inductance in Parallel</p>	 <p>$f_1 = \frac{1}{2\pi\sqrt{LC}}$</p>		$Z_o = \frac{RX_L X_C [X_L X_C - j(RX_L - RX_C)]}{(RX_L - RX_C)^2 + X_L^2 X_C^2}$ $ Z_o = \frac{RX_L X_C}{\sqrt{(RX_L - RX_C)^2 + X_L^2 X_C^2}}$ <p>= R when $X_L = X_C$ = 0 when $X_L = X_C$</p> $\phi = \tan^{-1} \frac{RX_L - RX_C}{X_L X_C}$

FIG. 32a. Equivalent impedances of parallel combinations of L , C , and R .

range of frequencies involved. Then, letting $Z_1 = Z_2 = R_0$, the loss is

$$L = 20 \log_{10} \left| \frac{2Z_s + R_0}{2Z_s} \right| \text{ db} = 20 \log_{10} \sqrt{1 + \frac{\cos \phi}{K} + \frac{1}{4K^2}} \text{ db} \tag{20a}$$

where $K = |Z_s|/R_0$ and ϕ is the phase angle of Z_s . For various values of K and ϕ the loss can be read from the curve (Fig. 34).

2. *Series Impedance.* The loss in decibels due to inserting a series impedance Z_s (Fig. 33a and c) is

$$L = 20 \log_{10} \left(\frac{Z_1 + Z_2 + Z_s}{Z_1 + Z_2} \right) \text{ db} \tag{21}$$

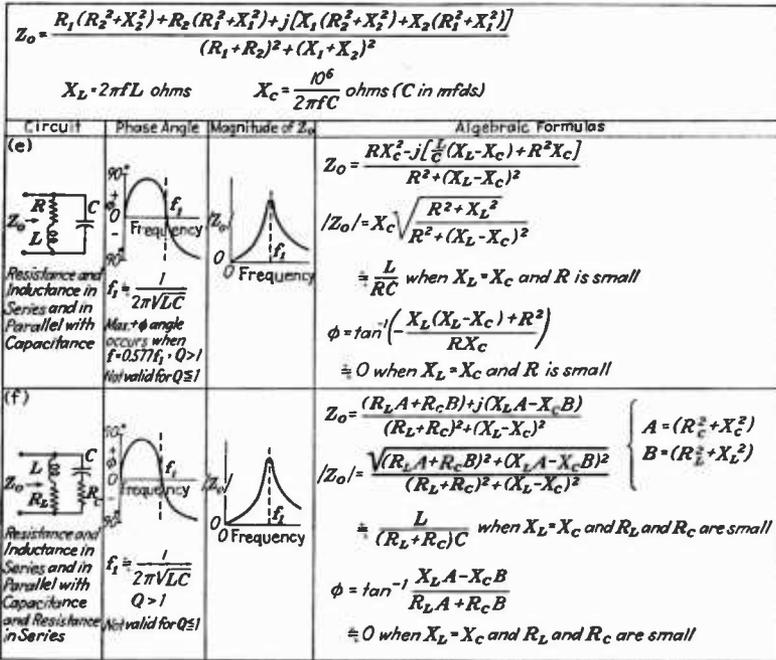


FIG. 32b. Equivalent impedances of parallel combinations of L, C, and R.

The series impedance can usually be inserted at a point in the circuit where the impedances Z_1 and Z_2 are matched and where each is substantially a pure resistance through the range of frequencies involved. Then, letting $Z_1 + Z_2 = R_o$, the loss is

$$L = 20 \log_{10} \left| \frac{R_o + Z_s}{R_o} \right| \text{ db}$$

$$= 20 \log_{10} \sqrt{1 + 2K \cos \phi + K^2} \text{ db} \tag{21a}$$

where $K = |Z_s|/R_o$ and ϕ is the phase angle of Z_s . The loss can be read from Fig. 34 for various values of K and ϕ .

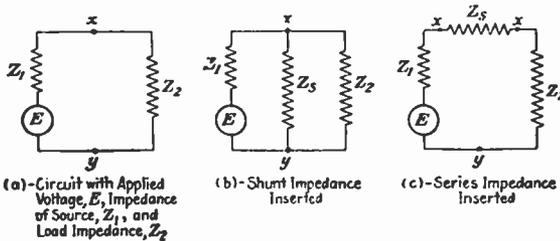


FIG. 33. Shunt and series impedances inserted in audio-frequency circuits.

14. Series Resonance.

Resonance frequency:

$$f_r = \frac{1}{2\pi\sqrt{LC}} \tag{22}$$

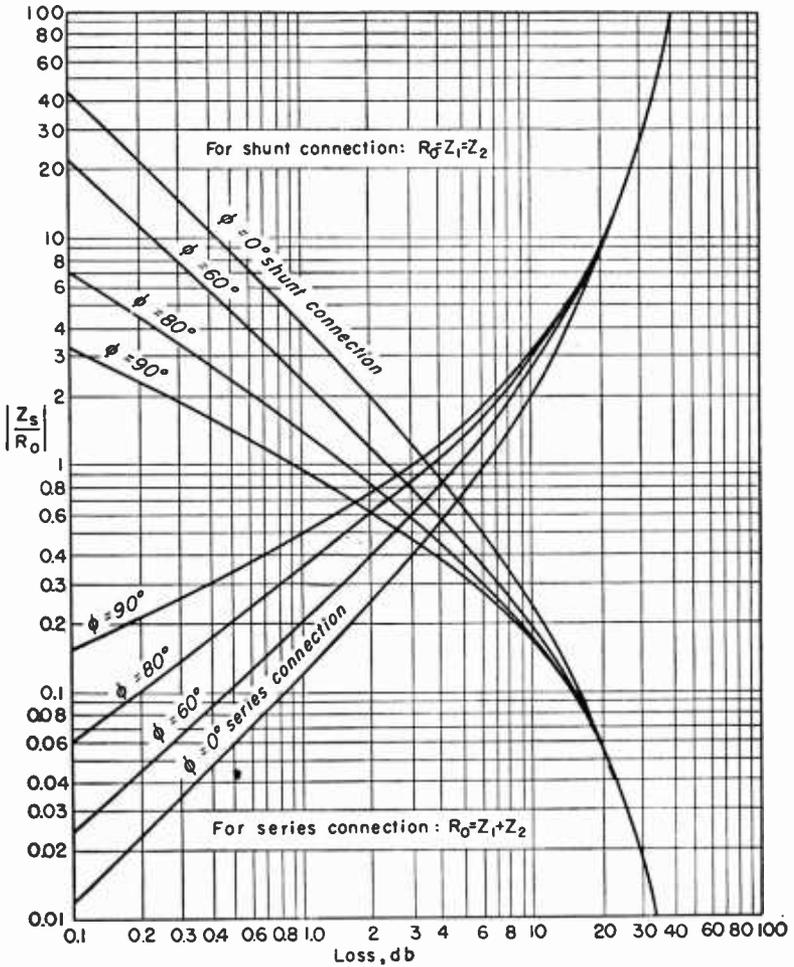
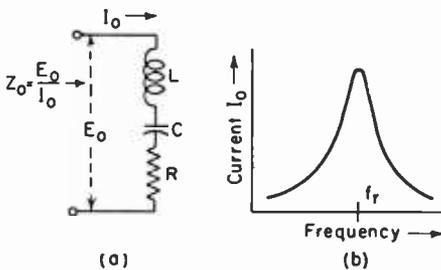


FIG. 34. Transmission loss due to insertion of shunt or series impedance.



Corresponding Units of L , C , and f_r for Eq. (22)

L	C	f_r
Henrys	Farads	Cycles
Henrys	μf	kc
μh	μf	Mc
μh	$\mu \mu f$	Mc $\times 10^3$

FIG. 35. (a) Series-resonant circuit and (b) current vs. frequency for constant E_0 .

Impedance at resonance:

$$Z_r = R_r \quad \text{ohms} \tag{23}$$

Impedance at any frequency, f_1 :

$$Z_1 = R_1 + j2\pi L \left(\frac{f_1^2 - f_r^2}{f_1} \right) = |Z_1| \angle \phi_1 \quad \text{ohms} \tag{24}$$

$$|Z_1| = \sqrt{R_1^2 + 4\pi^2 L^2 \left(\frac{f_1^2 - f_r^2}{f_1} \right)^2} \tag{25}$$

$$\phi_1 = \tan^{-1} \frac{X_1}{R_1} = \tan^{-1} \left[\frac{2\pi L(f_1^2 - f_r^2)}{f_1 R_1} \right] \tag{26}$$

Subscript_(r) denotes values at resonance, and subscript₍₁₎ values at any frequency f_1 .

Impedance in terms of Q :

Either of two conditions may be assumed: (1) Q is constant as frequency varies; (2) R is constant as frequency varies. Neither is exactly true in any practical case because R actually varies with the frequency in a complex manner. Expressions in terms of Q , however, give fair approximations if $Q = 10$ or more and are, therefore, useful.

The differences for impedance between constant R and constant Q are a few per cent in impedance near resonance for Q on the order of 10. For large Q and frequencies farther from resonance, the impedances by either assumption tend to become equal. The differences in phase-angle results are largest at frequencies farthest from resonance and for small Q 's. Near resonance and with large values of Q , the values by both assumptions approach equality.

For constant R :

$$R = \frac{\omega_r L}{Q_r} = \frac{\omega_1 L}{Q_1} \tag{27}$$

$$\frac{|Z_1|}{\omega_r L} = \sqrt{\frac{1}{Q_r^2} + n^2 + \frac{1}{n^2} - 2} \tag{28}$$

$$\phi_1 = \tan^{-1} \left[Q_r \left(n - \frac{1}{n} \right) \right] \tag{29}$$

where $n = f_1/f_r$.

For constant Q :

$$Q = \frac{\omega_r L}{R_r} = \frac{\omega_1 L}{R_1}$$

$$\frac{|Z_1|}{\omega_r L} = \sqrt{\frac{n^2}{Q^2} + n^2 + \frac{1}{n^2} - 2} \tag{30}$$

$$\phi_1 = \tan^{-1} \frac{Q}{n} \left(n - \frac{1}{n} \right) \tag{31}$$

where $n = f_1/f_r$.

Subscript_(r) denotes values at resonance, and subscript₍₁₎ values at any frequency f_1 .

At resonance, the voltage drop (E_L) across L is numerically equal to the voltage drop (E_C) across C . In terms of applied voltage E_0 when $Q > 25$,

$$|E_L| = |E_C| = QE_0$$

or

$$Q = \frac{|E_C|}{E_0} \tag{32}$$

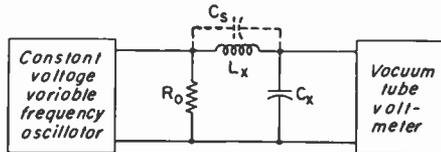


FIG. 36. Direct-reading Q meter using series resonance principle.

The latter equation is the basis of the direct-reading Q meter illustrated in Fig. 36. The inductance L_x to be measured is connected in series with a low-loss variable capacitor C_x . Voltage E_0 is applied from a constant-voltage variable-frequency oscil-

lator terminated in a very small resistance R_0 . The frequency and series capacitances are adjusted for series resonance. A high-input-impedance vacuum-tube voltmeter reads the series capacitor voltage $|E_C|$ and may be calibrated directly in terms of Q from Eq. (32), the input voltage being known and constant. In precision instru-

Product LC for Any Frequency, F

(1)		(2)			(3)		(4)		(5)			(6)
F $\mu\mu f$ - μh -kc $\mu\mu f$ -h-cps		LC			F μf -h-cps $\mu\mu f$ -h-kc		F $\mu\mu f$ - μh -kc $\mu\mu f$ -h-cps		LC			F μf -h-cps $\mu\mu f$ -h-kc
100	0	253	30	30	10		400	0	15	83	14	40
105	0	229	75	30	10	5	410	0	15	06	85	41
110	0	209	34	10	11		420	0	14	35	96	42
115	0	191	53	30	11	5	430	0	13	69	95	43
120	0	175	90	40	12		440	0	13	08	38	44
125	0	162	11	20	12	5	450	0	12	50	88	45
130	0	149	88	30	13		460	0	11	97	08	46
135	0	138	98	60	13	5	470	0	11	46	67	47
140	0	129	23	60	14		480	0	10	99	40	48
145	0	120	47	70	14	5	490	0	10	55	00	49
150	0	112	58	00	15		500	0	10	13	21	50
155	0	105	43	20	15	5	520	0	9	36	77	52
160	0	98	94	60	16		540	0	8	68	67	54
165	0	93	04	05	16	5	550	0	8	37	37	55
170	0	87	64	80	17		560	0	8	07	73	56
175	0	82	71	10	17	5	580	0	7	52	98	58
180	0	78	18	00	18		600	0	7	03	62	60
185	0	74	01	10	18	5	620	0	6	58	95	62
190	0	70	16	70	19		640	0	6	18	42	64
195	0	66	61	48	19	5	650	0	5	99	53	65
200	0	63	32	57	20		660	0	5	81	50	66
210	0	57	43	83	21		680	0	5	47	80	68
220	0	52	33	53	22		700	0	5	16	94	70
230	0	47	88	34	23		720	0	4	88	62	72
240	0	43	97	62	24		740	0	4	62	57	74
250	0	40	52	84	25		750	0	4	50	31	75
260	0	37	47	09	26		780	0	4	38	54	76
270	0	34	74	66	27		780	0	4	16	34	78
280	0	32	30	90	28		800	0	3	95	78	80
290	0	30	11	93	29		820	0	3	76	71	82
300	0	28	14	48	30		840	0	3	58	98	84
310	0	26	35	83	31		850	0	3	50	59	85
320	0	24	73	66	32		860	0	3	42	48	86
330	0	23	26	01	33		880	0	3	27	09	88
340	0	21	91	20	34		900	0	3	12	72	90
350	0	20	67	77	35		920	0	2	99	27	92
360	0	19	54	50	36		940	0	2	86	67	94
370	0	18	50	28	37		950	0	2	80	66	95
380	0	17	54	18	38		960	0	2	74	85	96
390	0	16	65	37	39		980	0	2	63	74	98
400	0	15	83	14	40		1000	0	2	53	30	100

NOTES: The double rules in Cols. 1 and 4 indicate that the decimal point in Cols. 2 and 5 is at the double rule; the single rules in Cols. 3 and 6 indicate that the decimal point in Cols. 2 and 5 is at the first single rule. Thus, the LC product for 100 cycles (Col. 1) is 2,533,030 (Col. 2) when the units are $\mu\mu f$ and h; or is 2.533030 (Col. 5) if the units are μf and h (Col. 6).

For frequencies not listed, linear interpolation may be used for accuracies of 0.1%. For higher accuracies, $LC = 253303/\beta$, neglecting decimals.

ments, correction factors are applied to eliminate the loading effects of R_0 and the voltmeter impedance, and of the distributed capacitance C_s of the coil.

15. Series-resonant Circuits as Equalizers. Series-resonant circuits are often used as equalizers where it is required to eliminate or attenuate a certain frequency or a small band of frequencies. The resonant circuit with a variable resistance in series is connected in shunt across the line or terminals of the circuit to be equalized, and more or less readily by-passes currents of the resonant and adjacent frequencies, depending upon the adjustment of R (see Fig. 37).

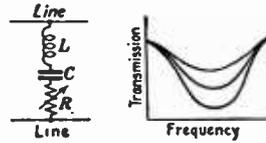


FIG. 37. Series-resonant equalizer.

16. Scratch Filters. A series-resonant circuit is the simplest form of scratch-and-hiss filter for electric phonographs. The resonance frequency is usually about 4,500 cycles; a typical filter is shown in Fig. 38 with its loss vs. frequency characteristic. A low-pass filter with 5,000-cycle cutoff is much better for the purpose, however (see Sec. 62).

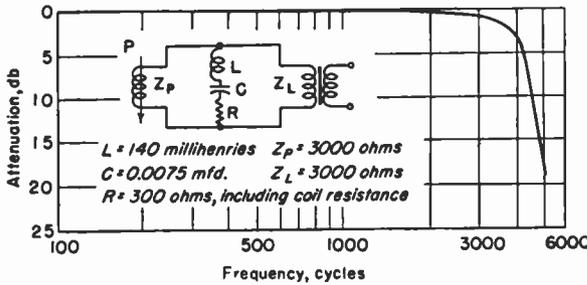


FIG. 38. Transmission characteristic of scratch filter used with magnetic phonograph pickup.

17. Tone Control. A series circuit resonant at about 1,000 cycles is sometimes used as tone control in an a-f amplifier. It may have a variable resistance and be connected in shunt in a grid or plate circuit, or it may be shunted across part of a volume control. Such a tone control tends to compensate automatically for the frequency characteristic of the ear, which varies with sound volume.

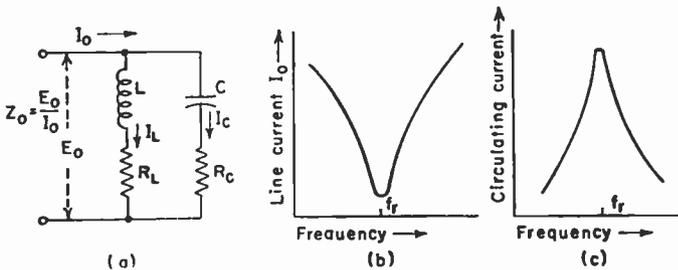


FIG. 39. Parallel-resonant circuit and characteristics for constant E_0 applied.

18. Parallel Resonance; Frequency Relations. There are various definitions of resonance in parallel LCR circuits, each of which gives a slightly different value of resonance frequency. The following are the most useful equations for parallel resonance frequency in radio applications.

In all cases R_L and R_C are in ohms, L in henrys, and C in farads, giving f_r in cycles.

1. *Free Oscillation Resonance* (R_C negligible).

$$f_r = \frac{1}{2\pi \sqrt{LC}} \sqrt{1 - \frac{R_L^2 C}{4L}} \quad (33)$$

This is the natural frequency of oscillation in a parallel circuit excited by a single pulse. (Sec Sec. 5 also.)

2. *L Variable and Adjusted for Maximum $|Z_0|^*$* (R_C negligible). If R_L is constant as L varies,

$$f_r = \frac{1}{2\pi \sqrt{LC}} \sqrt{1 + \frac{R_L^2 C}{L}} \quad (34)$$

If L/R_L is constant as L varies,

$$f_r = \frac{1}{2\pi \sqrt{LC}} \quad (35)$$

3. *Frequency Variable and Adjusted for Maximum $|Z_0|$* (R_C negligible).

$$f_r = \frac{1}{2\pi \sqrt{LC}} \sqrt{\sqrt{1 + \frac{2R_L^2 C}{L}} - \frac{R_L^2 C}{L}} \quad (36)$$

4. *C Variable and Adjusted for Maximum $|Z_0|$* (R_C negligible).

$$f_r = \frac{1}{2\pi \sqrt{LC}} \sqrt{1 - \frac{R_L^2 C}{L}} \quad (37)$$

In this case, the phase angle of Z_0 is zero at resonance ($Z_0 =$ pure resistance), and the line current I_0 is in phase with applied voltage E_0 .

The differences between these resonance frequencies in ordinary circuits are of the order of a fraction of 1 per cent and are trivial except when accurate measurements or extremely critical circuit adjustments are involved. An example of the latter case is the adjustment of a transmitter tank circuit for the highest possible efficiency. Condition 4, which gives zero phase angle and unity power factor for the equivalent impedance of the circuit at resonance, is the most common definition of parallel resonance.

5. *$Z_0 =$ Pure R , and R_C and R_L Not Negligible.*

$$f_r = \frac{1}{2\pi \sqrt{LC}} \sqrt{\frac{L - R_L^2 C}{L - R_C^2 C}} \quad (38)$$

6. *$Z_0 =$ Pure R , and R_C Negligible.*

$$f_r = \frac{1}{2\pi \sqrt{LC}} \sqrt{1 - \frac{R_L^2 C}{L}} = \frac{1}{2\pi \sqrt{LC}} \sqrt{1 - \frac{1}{Q_r^2}} \quad (\text{approx}) \quad (39)$$

where Q_r is the coil Q at resonance. This is the most useful expression for f_r for practical applications.

7. *Approximation When $R_L = R_C$ or When R_C and R_L Are Both Negligible.*

$$f_r = \frac{1}{2\pi \sqrt{LC}} \quad (40)$$

This ignores circuit resistance entirely and is used when approximate results are satisfactory.

* $|Z_0|$ is the magnitude of the equivalent impedance.

8. *Special Case When $R_L = R_C = \sqrt{L/C}$.* The circuit is resonant at all frequencies and $Z_0 = R_L = R_C = \sqrt{L/C}$ at all frequencies. Since L/C is usually of the order of 10^6 , R will be about 1,000 ohms, which limits the practical applications of the circuit.

19. *Impedance Relations in Parallel-resonant Circuits.* Exact impedance and phase-angle equations are given in Fig. 32*b*. The following are practical approximations.

Impedance at resonance (R_C negligible):

$$Z_r = \frac{X_C^2}{R_L} = \frac{X_L^2}{R_L} = \frac{X_L X_C}{R_L} = \frac{L}{R_L C} \quad \text{ohms (resistive)} \quad (41)$$

Impedance at any frequency, f_1 (R_C negligible):

$$Z_1 = X_C \frac{R_L X_C - j(R_L^2 + X_L^2 - X_L X_C)}{R_L^2 + (X_L - X_C)^2} = |Z_1|/\phi_1 \quad (42)$$

$$|Z_1| = \frac{X_C \sqrt{R_L^2 + X_L^2}}{\sqrt{R_L^2 + (X_L - X_C)^2}} \quad (43)$$

$$\phi_1 = \tan^{-1} \left(-\frac{R_L^2 + X_L^2 - X_L X_C}{R_L X_C} \right) \quad (44)$$

If $R_L \ll X_L$,

$$|Z_1| = \frac{L}{C} \frac{1}{\sqrt{R_L^2 + (X_L - X_C)^2}} = \frac{L}{C} \frac{1}{\sqrt{R_L^2 + 4\pi^2 L^2 \left(\frac{f_1^2 - f_r^2}{f_1}\right)^2}} \quad (45)$$

Subscript_(r) denotes values at resonance, and subscript_(1) values at any frequency, f_1 .

Impedance in terms of Q :

The comments in Sec. 14 relative to variation of R and Q in series-resonant circuits also apply here.

At resonance (R_C negligible, $Q > 25$):

$$Z_r = \frac{Q_r}{\omega_r C} = Q_r \omega_r L = R Q_r^2 \quad \text{ohms} \quad (46)$$

and

$$\frac{Z_r}{\omega_r L} = Q_r \quad (47)$$

where Q_r = coil Q at resonance.

At any frequency, f_1 (R_C negligible, $Q > 25$) for constant R :

$$\frac{|Z_1|}{\omega_r L} = \frac{\sqrt{\frac{1}{Q_r^2} + n^2}}{n \sqrt{\frac{1}{Q_r^2} + n^2 + \frac{1}{n^2} - 2}} \quad (48)$$

$$\phi_1 = \tan^{-1} [-n Q_r (n^2 - 1)] \quad (49)$$

where $n = f_1/f_r$.

Figure 40 is a family of resonance curves for $|Z_1|/\omega_r L$ vs. frequency ratio n , plotted for the condition of constant R . Figure 41 is a corresponding set of curves for ϕ_1 .

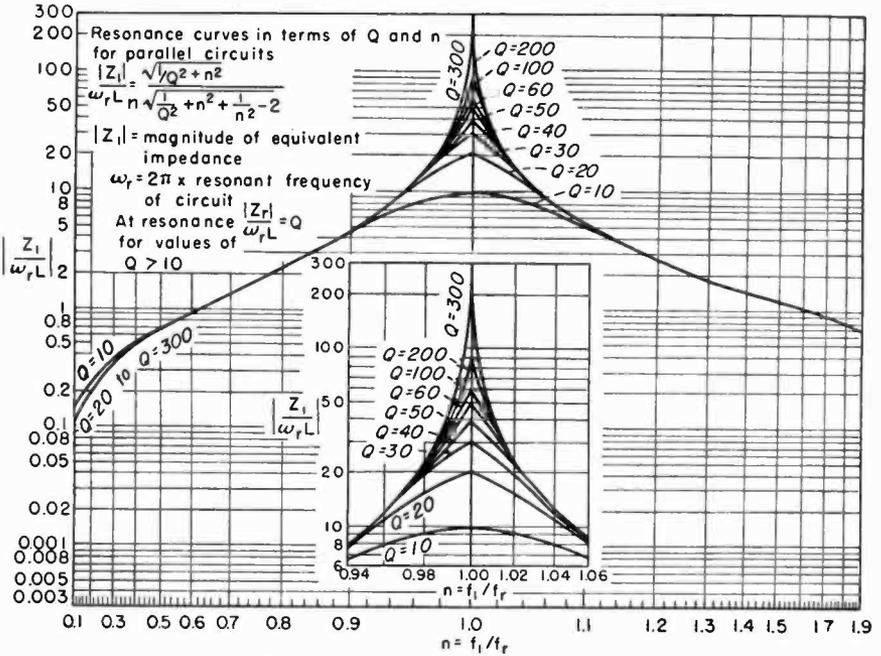


FIG. 40. Parallel-resonance curves.

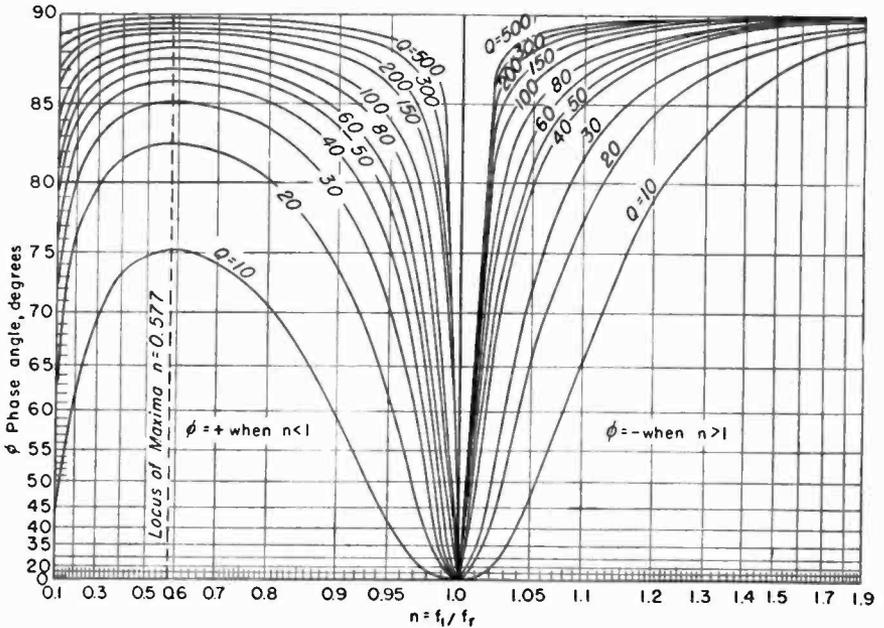


FIG. 41. Phase angle of parallel LC circuit in terms of n and Q .

For constant Q (R_C negligible, $Q > 25$):

$$\frac{|Z_1|}{\omega_r L} = \frac{\sqrt{(1/Q^2) + 1}}{\sqrt{(n^2/Q^2) + n^2 + (1/n^2) - 2}} \tag{50}$$

$$\phi_1 = \tan^{-1} [-Q(n^2 - 1)] \tag{51}$$

where $n = f_1/f_r$.

In all cases, ϕ_1 is positive when f_1 is less than f_r and is negative when f_1 is greater than f_r .

When the circuit Q is 25 or larger, the impedance of a parallel LCR circuit at $f_1 = 0.707f_r$ and at $f_1 = 1.414f_r$ is

$$|Z_1| = \frac{1.414}{Q_r} |Z_r| \quad \text{or} \quad \frac{|Z_1|}{|Z_r|} = \frac{1.414}{Q_r} \tag{52}$$

This is the basis for one method of measuring the Q of an inductance coil. The coil is connected in parallel with a low-loss variable capacitor, and voltage is applied from a constant-voltage variable-frequency source. The line current (I_r) is measured at resonance frequency and at 70.7 per cent or 141.4 per cent of resonance frequency (I_1), and Q is then determined from the relation

$$Q = 1.414 \frac{|I_1|}{|I_r|}$$

20. Line and Circulating Currents in Parallel-resonant Circuits. The true circulating current in a parallel-resonant circuit operating at or very near resonance frequency is the current representing the energy oscillating to and fro between L and C . This is a relatively large current (as compared to line current) and is a component of both the L branch and the C branch currents I_L and I_C (see Fig. 39). A second component of I_L is the current representing the losses in the L branch, which include the I^2R_L loss due to the equivalent ohmic resistance of the branch plus the energy taken from the circuit by any inductively coupled load. A second component of I_C represents ohmic losses in the C branch plus any load that may be coupled to that branch.

In addition, if the applied voltage E_0 is not sinusoidal and of frequency $= f_r$, the harmonic currents through the L and C branches will not be equal because the reactances of the two branches are equal (approximately) only at the resonance frequency f_r . Thus currents produced by components of E_0 whose frequency is less than f_r will flow principally through the L branch, while those whose frequency is greater than f_r will flow principally through the C branch. Hence the total I_L is not actually equal to the total I_C and the difference is the line current, I_0 .

In some cases, when the total losses of the LCR circuit are small, it is assumed that the circulating current is $I_L = I_C$. At the same time the line current cannot be ignored but is given by the approximation

$$I_0 = \frac{I_L}{Q} = \frac{I_C}{Q} \tag{53}$$

or, in terms of I_0 ,

$$I_L = I_C = QI_0 \tag{53a}$$

This approximation is most nearly accurate when the applied voltage is sinusoidal and has frequency f_r and when the circuit Q is high; it is inaccurate under any other conditions.

21. Design of Parallel-resonant Circuits. To design a parallel-resonant circuit, we have to determine values of L , C , R_L , and R_C to satisfy a given set of conditions. Values of $|Z_r|$ at resonance, the resonance frequency, and Q will first have to be determined by analysis of the intended use of the resonant circuit.

In a vacuum-tube oscillator, for example, f_r of the tuned circuit is known, and $|Z_r|$ at resonance is fixed by the permissible plate voltage swing. For Q (which includes the effect of the external load coupled to the tuned circuit, as well as the latter's ohmic resistance) a value of from 12 to 20 represents a good compromise between oscillator efficiency and frequency stability.

Another example of the factors involved in the choice of Q in an application is that of a tuned circuit for an r-f amplifier to pass a modulated carrier. In this case the LC circuit must have sufficient decrement to damp out its own natural oscillations between successive peaks of modulation; otherwise there is an effective decrease in modulation percentage with a corresponding loss of fidelity. If the carrier frequency is f_c and the modulation frequency f_m , the maximum decrement of the modulated carrier wave at 100 per cent modulation is approximately

$$\delta_1 = 2.303 \log_{10} \left(\frac{1}{1 - \pi \frac{f_m}{f_c}} \right) \quad (54)$$

The decrement δ_2 of the tuned circuit should be 10 to 20 times as large as δ_1 for faithful response. Then Q for the tuned circuit is

$$Q = \frac{\pi}{\delta_2} \quad (55)$$

The value of $|Z_r|$ at resonance will depend upon plate-load impedance requirements of the amplifier tube.

In some cases the ratio of volt-amperes circulating in LC to watts dissipated is the basis for the design of an LC circuit; in this case

$$Q = \frac{\text{volt-amp stored}}{\text{watts dissipated}} \quad (56)$$

The effect of any load coupled to a tuned circuit must be taken into account as part of the total effective R of the circuit. If the power taken by the load is W_d watts and I_C is the circulating current in LCR , the total equivalent impedance of the circuit is, approximately,

$$R = R_0 + \frac{W_d}{I_C^2} \quad \text{ohms} \quad (\text{approx}) \quad (57)$$

where R_0 is the ohmic resistance.

Examples of Design of Parallel-resonant Circuit. Assume that a parallel circuit (Fig. 39) is to be resonant at 5,000 cycles, with an impedance of 4,000 ohms at resonance ($n = 1$) and an impedance of 100 ohms at 3,000 cycles ($n = 0.6$). From Fig. 40, $|Z_r|/\omega_r L = 0.9$ for all values of Q when $n = 0.6$. At resonance $|Z_r|/\omega_r L$ is to be $(4,000/100) \times 0.9 = 36$. From the curves it is found that $Q = 36$ gives $|Z_r|/\omega_r L = 36$ at $n = 1$ where $\omega_r = 31,416$. Then for $n = 1$,

$$i, \quad = 36\omega_r L = 4,000 \quad \text{or} \quad L = \frac{4,000}{36 \times 31,416} = 0.00354 \quad \text{henry}$$

LC for 5,000 cycles = 10.136×10^{-10} . Then $C = LC/L = 0.286 \times 10^{-6}$ farad, and $R = \omega_r L/Q = 3.08$ ohms.

As a second example suppose there is to be designed a tuned circuit for an r-f amplifier which requires a plate-load impedance of 10,000 ohms and which is to amplify a 1,000-kc carrier with amplitude modulation up to 5,000 cycles. From Eq. (54),

$$\delta_1 = 2.303 \log_{10} \frac{1}{1 - \pi \frac{5 \times 10^3}{10^6}} = 0.0159$$

$$\delta_2 = 20 \delta_1 = 0.318$$

$$Q = \frac{\pi}{\delta_2} = 9.85 = \frac{|Z_r|}{\omega_r L} \quad \text{from Eq. (55)}$$

$$\omega_r L = \frac{|Z_r|}{Q} = \frac{10,000}{9.85} = 1,015$$

$$L = \frac{\omega_r L}{\omega_r} = \frac{1,015}{2\pi f_c} = 162 \quad \mu\text{h}$$

$$LC \text{ for } 1,000 \text{ kc} = 2.53 \times 10^{-10}$$

$$C = \frac{LC}{L} = \frac{2.53 \times 10^{-10}}{162 \times 10^{-6}} = 157 \quad \mu\mu\text{f}$$

$$R = \frac{\omega_r L}{Q} = \frac{1,015}{9.85} = 103 \quad \text{ohms}$$

This consists of the ohmic resistance of LCR plus the equivalent R of the coupled load, as computed by Eq. (57).

22. Design of Oscillator Tracking Circuits. In superheterodyne receivers with "ganged" capacitor tuning, the most common method for tracking the oscillator-tuned circuit at a constant frequency difference from the r-f circuits is by means of a gang capacitor with identical sections and with an adjustable padding capacitor in series with the oscillator section. A typical oscillator circuit of this type is shown in Fig. 42. The tracking is approximate as no combination of C₁, C₂, and C₃ will give

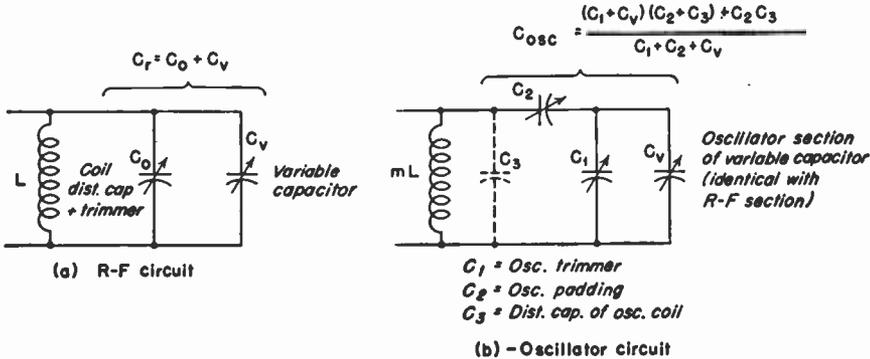


FIG. 42. Oscillator tracking circuits for superheterodynes.

perfect alignment at more than three points on the dial as shown in Fig. 43. These points are usually chosen near the ends and the middle of the frequency range—in a broadcast receiver, for example, at 1,400, 1,000, and 600 kc. Slight tracking errors will exist at all other frequencies in the band; these will be approximately proportional to the i-f frequency used. The maximum errors are at the ends of the band and amount to about 2 kc for a good design with an i.f. of 175 kc.

The values of C₁, C₂, and C₃ may be determined by calculation or by experimental methods. Either method involves a considerable amount of labor. The following design procedure, due to Roder,¹ is probably the most direct method of solution (six-place logarithms or a calculating machine are recommended for all calculations):

Step 1. Known constants:

- a. Three frequencies of perfect alignment (= f). (Usually 1,400, 1,000, and 600 kc for broadcast receivers.)
- b. R-f circuit inductance (= L).
- c. R-f circuit trimmer capacitance (= C₀). (Including distributed capacitance of r-f coil.)

¹ Roder, Hans, Oscillator Padding, *Radio Eng.*, March, 1935, p. 7.

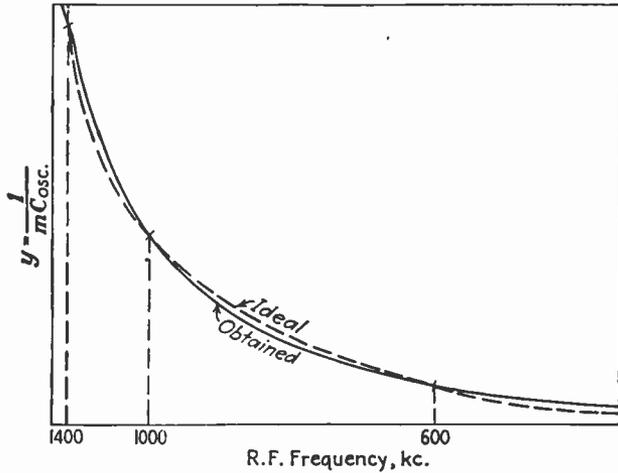


FIG. 43. Closeness of tracking secured by formulas.

- d. Intermediate frequency ($= f_i$).
 e. Distributed capacitance of oscillator coil ($= C_1$).

Solution to yield: Values of C_1 , C_2 , and L_{osc} .

Units: All constants are measured in the following units:

- f = frequency, kc
 L = inductance, μh
 C = capacitance, $\mu\mu f$

Step 2. Compute

$$x_n = \frac{253.3 \times 10^8}{L f_i^2} \quad \text{and} \quad y_n = \frac{f_{osc}^2 L}{253.3 \times 10^8}$$

for each alignment frequency.

Step 3. Compute

$$X = \frac{y_2 - y_1 + x_2 B - x_1 A}{B - A} \quad Y = \frac{y_1 B - y_2 A}{B - A}$$

where $A = \frac{y_1 - y_2}{x_2 - x_1}$

$$B = \frac{y_2 - y_1}{x_1 - x_2}$$

Step 4. Compute

$$K = (x_1 - X)(y_1 - Y) \equiv (x_2 - X)(y_2 - Y) \equiv (x_3 - X)(y_3 - Y)$$

(The truth of these identities is a check on the accuracy of the calculations thus far.)

Step 5. Compute

$$m = \frac{1}{K} (1 - v)$$

where $v = 0.5u - 0.3125u^2 + 0.2188u^3$

$$u = \frac{4C_1 Y}{K}$$

Step 6. Compute

$$C_1 = C_0 - X - \frac{Kv}{2Y} (1 + 0.75v + 0.625v^2 + 0.547v^3)$$

and

$$C_2 = \frac{1}{Y} \sqrt{\frac{K}{m}}$$

Step 7. Compute $L_{osc} = mL$

Example:

Step 1. Let

$$\begin{aligned} f_{r_1} &= 1,400 \text{ kc} & L &= 200\mu\text{h} \\ f_{r_2} &= 1,000 \text{ kc} & C_0 &= 30 \mu\mu\text{f} \\ f_{r_3} &= 600 \text{ kc} & C_2 &= 15\mu\mu\text{f} \\ f_i &= 175 \text{ kc} \end{aligned}$$

Step 2.

	f_r	f_{osc}	r	v
1	$1,400 \times 10^3$	$1,575 \times 10^3$	64.617	0.0195865
2	$1,000 \times 10^3$	$1,175 \times 10^3$	126.650	0.0109011
3	600×10^3	775×10^3	351.806	0.0047424

Step 3. $A = 140.0126 \times 10^{-6}$; $B = 27.3531 \times 10^{-6}$
 $X = -5.1103$ $Y = 0.1138 \times 10^{-3}$

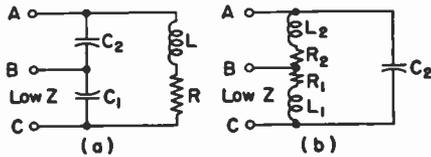
Step 4. $K = 1.2863$.

Step 5. $u = 5.3097$; $v = 2.5700 \times 10^{-2}$; $m = 0.7574$.

Step 6. $C_1 = 20.17 \mu\mu\text{f}$; $C_2 = 1451.5 \mu\mu\text{f}$.

Step 7. $L_{osc} = 151.50 \mu\text{h}$.

23. Tapped Tank Circuits. In some cases the high impedance of a parallel LCR circuit at resonance is a disadvantage, e.g., at the end of a low-impedance transmission line where the correct termination is about 500 ohms. However, the low impedance can be obtained by tapping the LCR circuit in either the L or C branch as shown in Fig. 44. The result is a coupled circuit, that part of the reactance between B and C being the mutual impedance.



Capacitance tapped Inductance tapped

FIG. 44. Tapped tank circuits.

1. *Capacitance Tapped.* In Fig. 44a, the impedance at B-C is

$$|Z_{BC}| = \frac{\sqrt{L^2 C_2^2 \left(\frac{1}{C_1(C_1 + C_2)} \right)^2 + \frac{R^2 L C_2}{C_1(C_1 + C_2)}}}{R} \tag{58}$$

If R is small,

$$|Z_{BC}| = \frac{LC_2}{RC_1(C_1 + C_2)} \tag{58a}$$

and its ratio to the impedance Z_{AC} is

$$\frac{|Z_{BC}|}{|Z_{AC}|} = \frac{C_2^2}{(C_1 + C_2)^2} \tag{58b}$$

The resonant frequency is

$$f_r = \frac{1}{2\pi \sqrt{L \frac{C_1 C_2}{C_1 + C_2}}} \quad (59)$$

and the impedances Z_{AC} and Z_{BC} are both purely resistive at resonance.

The ratio of C_1 to C_2 for a given ratio between Z_{AC} and Z_{BC} is

$$\frac{C_1}{C_2} = \left(\sqrt{\frac{Z_{AC}}{Z_{BC}}} - 1 \right) \quad (60)$$

In terms of the resonant frequency, inductance, and the impedance ratio,

$$C_1 = \frac{1}{4\pi^2 f_r^2 L} \sqrt{\frac{Z_{AC}}{Z_{BC}}} \quad (61)$$

$$C_2 = \frac{1}{4\pi^2 f_r^2 L \left(1 - \sqrt{\frac{Z_{BC}}{Z_{AC}}} \right)} \quad (62)$$

2. *Inductance Tapped.* In Fig. 44b the inductance is tapped, and the impedance at B-C is (assuming no mutual inductance between L_1 and L_2)

$$|Z_{BC}| = \frac{\sqrt{\left(R_1 R_2 - \frac{L_1 L_2}{(L_1 - L_2) C_2} + \frac{L_2^2}{C_2} \right)^2 + \left(\frac{R_2 L_1}{\sqrt{(L_1 + L_2) C_2}} \right)^2}}{R_1 + R_2} + \frac{\frac{R_1 L_2}{\sqrt{(L_1 + L_2) C_2}} - \frac{R_1 \sqrt{(L_1 + L_2) C_2}}{C_2}}{C_2}} \quad (63)$$

If R_1 and R_2 are small,

$$|Z_{BC}| = \frac{L_2^2}{C_2 (R_1 + R_2) (L_1 + L_2)} \quad (64)$$

and its ratio to the total impedance Z_{AC} is

$$\frac{|Z_{BC}|}{|Z_{AC}|} = \frac{L_2^2}{(L_1 + L_2)^2} \quad (65)$$

The resonant frequency is

$$f_r = \frac{1}{2\pi \sqrt{(L_1 + L_2) C_2}} \quad (66)$$

and the impedances Z_{AC} and Z_{BC} are both resistive at resonance.

The ratio of L_1 to L_2 for a given ratio between Z_{AC} and Z_{BC} is

$$\frac{L_1}{L_2} = \sqrt{\frac{Z_{AC}}{Z_{BC}}} - 1 \quad (67)$$

In terms of the frequency, capacitance, and the impedance ratio,

$$L_1 = \frac{1}{4\pi^2 f_r^2 C_2} \left(1 - \sqrt{\frac{Z_{BC}}{Z_{AC}}} \right) \quad (68)$$

$$L_2 = \frac{1}{4\pi^2 f_r^2 C_2} \sqrt{\frac{Z_{BC}}{Z_{AC}}} \quad (69)$$

COUPLED CIRCUITS

24. **Coupling.** Two circuits are coupled if they have a common impedance so situated that it causes the current flowing in one circuit to influence the current in the other. The common impedance may be resistance, capacitance, inductance, mutual inductance, or any combination of these. In the general case where it contains both resistance and reactance, the analysis is complicated.¹

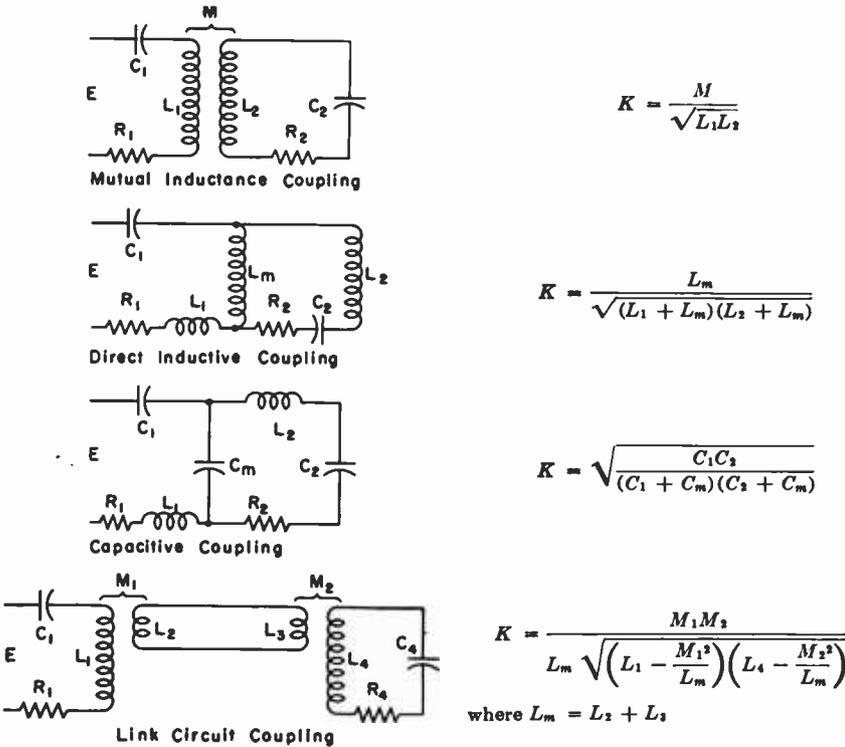


FIG. 45. Commonly used types of coupled circuits and their coupling coefficients.

25. **Coefficient of Coupling.** The extent of coupling between two circuits is called the *coefficient of coupling* and is

$$K = \frac{X_m}{\sqrt{X_1 X_2}} \tag{70}$$

where X_m is any one component of the mutual impedance (ωL_m , $1/\omega C_m$, or ωM in ohms) and X_1 and X_2 are the total impedance components of the same kind in the respective circuits, including the mutual impedance. K varies in value between 0 and 1; if it is nearly 1, the coupling is *close* or *tight*; if it is near 0 ($K = 0.1$ or less), the coupling is *loose*.

26. **Coupling Coefficients for R-f Circuits.** Values of K for four types of coupling most used in r-f circuits are given in Fig. 45.

¹ For discussions of this case, see G. W. Pierce, "Electric Oscillations and Electric Waves," McGraw-Hill, 1920, and C. B. Aiken, Two-mesh Tuned Coupled Filters, *Proc. IRE*, February, 1937, p. 230.

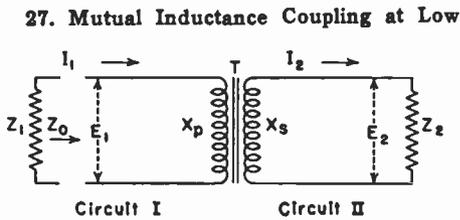


FIG. 46. Transformer coupling of low-frequency circuits.

Under these conditions, the following approximate relations hold:

Impedance looking into primary:

$$Z_0 = \frac{X_p}{X_s} Z_2 = \left(\frac{N_p}{N_s} \right)^2 Z_2$$

where N_p and N_s are number of turns in primary and secondary windings, respectively.

Current ratio:

$$\frac{I_1}{I_2} = \frac{N_s}{N_p} \quad I_2 = \frac{N_p}{N_s} I_1 \quad I_1 = \frac{N_s}{N_p} I_2$$

Voltage ratio:

$$\frac{E_1}{E_2} = \frac{N_p}{N_s} \quad E_1 = \frac{N_p}{N_s} E_2 \quad E_2 = \frac{N_s}{N_p} E_1$$

Volt-amperes:

$$E_1 I_1 = E_2 I_2 + \text{transformer losses}$$

Impedance ratio:

$$\frac{Z_0}{Z_2} = \left(\frac{N_p}{N_s} \right)^2 \quad \frac{Z_2}{Z_0} = \left(\frac{N_s}{N_p} \right)^2$$

These simple relations are not quite exact for any actual transformer because of core losses, flux leakage, and resistance and distributed capacitance of windings. Also, since these imperfections vary with frequency, the various ratios likewise change with frequency. A well-designed transformer may give results approaching 90 to 95 per cent of the values given.

To minimize the effects of these imperfections, transformer designs must take into account not only the ratios but also the actual values of primary and secondary currents and impedances. For this reason, a-f transformers are rated in terms of the actual impedances between which they are designed to operate, and the frequency range over which they will give satisfactory performance. Their performance outside these ratings is likely to be deficient. Thus, an a-f transformer designed to match one impedance of 500 ohms to another of 50 ohms (Z ratio of 10:1) probably would not give satisfactory results if used between impedances of 50,000 and 5,000 ohms, even though the impedance ratio is the same.

28. Mutual Inductance Coupling at Radio Frequencies. Excessive core losses, high resistance of windings, and large distributed capacitance bar the use of a-f types of iron-cored transformers in r-f circuits. Without iron cores, only small values of coupling between primary and secondary can be realized. Open cores of powdered iron are frequently used in r-f transformers, but their effect is principally to increase the circuit Q with little effect on the coupling. However, the range of frequencies that an r-f transformer is required to pass efficiently is usually small in comparison with the center frequency. The side bands of an a-f modulated signal, for instance, occupy a frequency band whose width is only 2 or 3 per cent of the carrier frequency.

(This is not true for all types of radio signals, however. A television signal may require a band width of 10 to 15 per cent of the carrier frequency. In this case, simple transformer circuits are not suitable coupling devices.)

When the band width is a small percentage of the center frequency, the limited values of K in an r-f transformer can be compensated by resonating the primary and secondary circuits. The transfer of energy in such an arrangement will vary with frequency, but this can be made reasonably constant over a narrow band that is still sufficiently wide to accommodate side bands.

29. Practical Relations in Transformer-coupled R-f Circuits. A practical application of a tuned coupled r-f circuit is shown in Fig. 47, which represents an i-f trans-

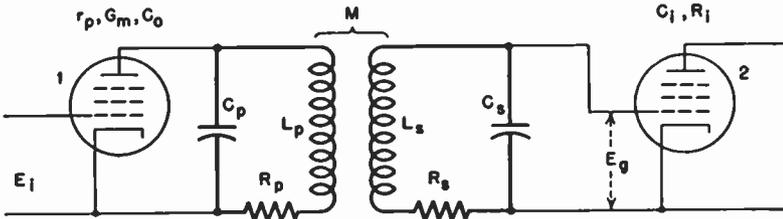


FIG. 47. R-f amplifier stage with tuned coupled circuits.

former used to couple the output of a pentode tube to the grid of the following tube. The same circuit and the formulas pertaining to it apply equally well to t-r-f coils with untuned primaries if C_p is eliminated from both circuit and equations.

For practical purposes, the effectiveness of the coupled circuit of Fig. 47 is determined by its gain and its selectivity. These quantities are dependent not only upon the L , R , and C of the coupling circuits themselves, but also upon the effective input and output impedances of the two tubes, which shunt the primary and secondary and alter their resistance and capacitance.

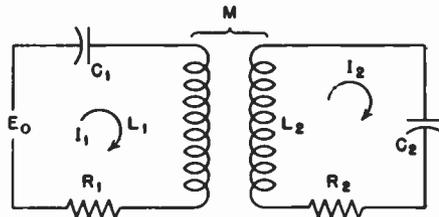


FIG. 48. Coupling circuit equivalent to that of Fig. 47.

The circuit of Fig. 47 is very nearly equivalent to that of Fig. 48 if

$$C_1 = C_p + C_o \quad L_1 = L_p \quad R_1 = \frac{1}{\omega^2 C_1^2 r_p} + r_p$$

$$C_2 = C_s + C_i \quad L_2 = L_s \quad R_2 = \frac{1}{\omega^2 C_2^2 R_i} + R_i$$

where C_o = output capacitance of tubes

C_i = input capacitance of tubes

r_p = tube plate resistance

R_i = tube input ohmic resistance

L_1 and L_2 = self-inductances of primary and secondary coils, each determined with the opposite circuit removed or uncoupled

For pentode tubes, the circuit input voltage is approximately

$$E_o = \frac{G_m}{\omega C_2} \times E_i \tag{71}$$

where E_i = input voltage to grid of first tube. The equivalent voltage E_o may be considered as coming from a source having zero internal impedance.

For triodes, the equivalent voltage is approximately

$$E_0 = \mu \times E_i \tag{72}$$

but in this case it must be considered as coming from a source whose internal impedance is equal to r_p , the plate resistance of the first tube. E_0 is the actual input voltage to the grid of the second tube.

30. Band-pass Characteristics of Coupled Circuits. If the primary and secondary circuits are individually tuned to the same frequency f_0 so that $L_1C_1 = L_2C_2$ (L_1 not necessarily equal to L_2 , nor C_1 to C_2), the secondary current I_2 varies with frequency according to one of the resonance curves of Fig. 49. Since E_0 applied to the grid of the following tube is nearly proportional to I_2 for narrow bands, these curves also represent output voltage and, therefore, the gain vs. frequency characteristic of the amplifier stage.

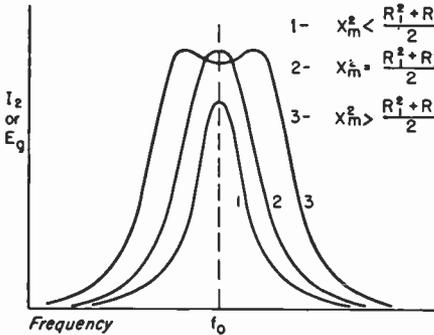


FIG. 49. Variation of secondary current and gain with frequency and degree of coupling.

CRITICAL COUPLING: When $X_m^2 = R_1R_2$, the coupling is called *critical*, and the curve has its maximum at f_0 , as shown by curve 2, Fig. 49. Critical coupling also gives the maximum possible peak value, and this is

$$I_{2 \max} = \frac{E_0}{2 \sqrt{R_1R_2}} \tag{approx} \tag{73}$$

or

$$E_0 \max = \frac{E_0}{2\omega_0C_2 \sqrt{R_1R_2}} \tag{approx} \tag{74}$$

where $\omega_0 = 2\pi f_0 = 1/\sqrt{L_1C_1} = 1/\sqrt{L_2C_2}$, and other quantities refer to the equivalent circuit of Fig. 48.

In terms of K and Q , the coupling is critical when $K = 1/\sqrt{Q_1Q_2}$, where Q_1 and Q_2 are the respective Q 's of the primary and secondary circuits. Practically, these are approximately the Q 's of the primary and secondary coils. If $Q_1 = Q_2 = Q$, $K = 1/Q$ for critical coupling.

UNDERCOUPLING: When $X_m^2 < R_1R_2$, the circuits are *undercoupled*. In this case, the curve has a lower maximum than for the case of critical coupling, as shown by curve 1, Fig. 49.

OVERCOUPLING: When $X_m^2 > R_1R_2$, the circuits are *overcoupled*. This may produce either a single or double peak, the latter case being shown by curve 3. In either case, the peak or peaks are equal to the maximum for critical coupling if $R_1 = R_2$, and is slightly less if $R_1 \neq R_2$.

2. Frequency Characteristic and Shape of Curves.

SINGLE PEAK CURVE: When $X_m^2 \leq (R_1^2 + R_2^2)/2$, the curve has a single peak (curves 1 and 2, Fig. 49) and has approximately the same shape as the resonance curve of a parallel LCR circuit whose $Q = \sqrt{Q_1Q_2}$.

* This was pointed out by Aiken, *loc. cit.*

DOUBLE PEAK CURVE: When $X_m^2 > (R_1^2 + R_2^2)/2$, the curve has a double peak (curve 3). If, in addition, $R_1 = R_2 = R$, the value of coupling for transition from the condition of single peak to double peak is the same as the value for critical coupling, and $X_m = R$. Furthermore, the peaks have maximum possible value.

3. Selectivity and Gain. Increasing coupling increases gain, up to the point of critical coupling, and reduces selectivity and, in practice, the best value of X_m is a compromise between gain and selectivity. For example, in i-f transformers for a-f modulated signals, it is common practice to adjust coupling to about 80 per cent of critical.

In i-f circuits where high fidelity of response is of major importance, over-coupling is frequently used to broaden the nose of the curve by double peaking. This passes efficiently the upper side bands due to higher a-f modulating frequencies. Such a curve is shown in curve 3 of Fig. 49 and is reproduced in Fig. 50 with dimensions added for reference in Sec. 31.

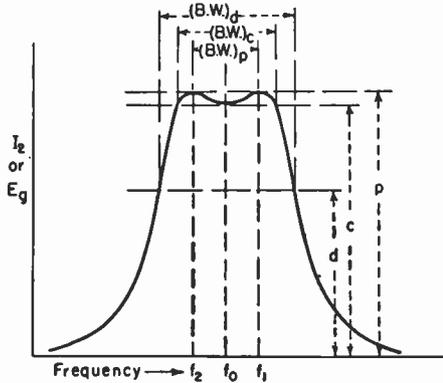


FIG. 50. Band-pass characteristic of over-coupled circuit.

31. Equations for Band Width and Gain. Although it is ordinarily easier to measure the response curve of a coupled circuit than it is to measure R_1 and R_2 and calculate the curve, relations between circuit parameters and gain and response are useful for design and analysis. The following relations give close approximations to exact values for the circuits of Figs. 47 and 48 and are based partly on Aiken's¹ results. The principal assumptions are that the band width is negligible compared to f_0 , that R_1 and R_2 are constant throughout the band width, and that E_i is proportional to I_1 .

Symbols used in gain and band-width equations:

- E_i = input voltage to first tube grid (Fig. 47)
- E_o = input voltage to second tube grid
- $\omega_0 = \frac{1}{\sqrt{L_1 C_1}} = \frac{1}{\sqrt{L_2 C_2}}$, both circuits tuned to $f_0 = \frac{\omega_0}{2\pi}$
- $K = \frac{M}{\sqrt{L_1 L_2}}$
- $Q_1 = \frac{\omega_0 L_1}{R_1}$
- $Q_2 = \frac{\omega_0 L_2}{R_2}$
- G_m = tube transconductance

Gain at center frequency for any value of coupling is proportional to dimension c at f_0 of Fig. 50 for either single or double peak curves and is

$$\frac{E_o}{E_i} = \frac{G_m M}{\omega_0 C_1 C_2 (R_1 R_2 + \omega_0^2 M^2)} = \frac{G_m \omega_0 K \sqrt{L_1 L_2}}{\frac{1}{Q_1 Q_2} + K^2} \tag{75}$$

¹ *Ibid.*

If both circuits are identical and $R_1 = R_2 = R$, $Q_1 = Q_2 = Q$, and $L_1 = L_2 = L$, the gain is

$$\frac{E_o}{E_i} = \frac{G_m M}{\omega_0 C_1 C_2 (R^2 + \omega_0^2 M^2)} = \frac{G_m \omega_0 K L}{\frac{1}{Q^2} + K^2} \quad (76)$$

Maximum possible gain at center frequency is proportional to dimension p at f_0 and occurs when coupling is critical ($\omega_0 M = R_1 R_2$ or $K = 1/\sqrt{Q_1 Q_2}$) for either single or double peak curves and is

$$\frac{E_o}{E_i} \max = \frac{G_m \omega_0^2 L_1 L_2}{2 \sqrt{R_1 R_2}} = \frac{G_m Q_1 Q_2 \sqrt{R_1 R_2}}{2} \quad (77)$$

If both circuits are identical and $R_1 = R_2 = R$, $Q_1 = Q_2 = Q$, and $L_1 = L_2 = L$, the gain is

$$\frac{E_o}{E_i} \max = \frac{G_m \omega_0^2 L^2}{2R} = \frac{G_m Q^2 R}{2} = \frac{G_m Q \omega_0 L}{2} \quad (78)$$

Gain at peaks of double peak curve (when $2\omega_0^2 M^2 > R_1^2 + R_2^2$) is the dimension p at f_1 and f_2 of Fig. 50 and is

$$\frac{E_o}{E_i} = \frac{2G_m M}{\omega_0 C_1 C_2 \sqrt{4\omega_0^2 M^2 (R_1 + R_2)^2 - (R_1^2 - R_2^2)}} \quad (79)$$

If $R_1 = R_2 = R$, the gain is

$$\frac{E_o}{E_i} = \frac{G_m}{2\omega_0^2 C_1 C_2 R} = \frac{Q_1 Q_2 R G_m}{2} = \frac{Q_1 \omega_0 L_1 C_m}{2} \quad (80)$$

Transition from single to double peak occurs when

$$f_0 = \frac{1}{2\pi M} \sqrt{\frac{R_1^2 + R_2^2}{2}} \quad \text{or} \quad M = \frac{1}{\omega_0} \sqrt{\frac{R_1^2 + R_2^2}{2}} \quad (81)$$

If $R_1 = R_2 = R$, transition occurs when

$$f_0 = \frac{R}{2\pi M} \quad \text{or} \quad M = \frac{R}{\omega_0} \quad (82)$$

Band width, peak to peak, when $L_1 = L_2 = L$ is the dimension $(BW)_p$ in Fig. 50 and is

$$(BW)_p = \frac{0.1592}{L} \sqrt{\omega_0^2 M^2 - \frac{R_1^2 + R_2^2}{2}} = f_0 \sqrt{K^2 - \frac{1}{2} \left(\frac{1}{Q_1^2} + \frac{1}{Q_2^2} \right)} \quad (83)$$

If $R_1 = R_2 = R$,

$$(BW)_p = \frac{0.1592}{L} \sqrt{\omega_0^2 M^2 - R^2} = f_0 \sqrt{K^2 - \frac{1}{4Q^2}} \quad (84)$$

If Q_1 and Q_2 are both large,

$$(BW)_p = K f_0 \quad (\text{approx}) \quad (85)$$

Band width between points on curve where gain equals gain at center frequency is $(BW)_c$ in Fig. 50 and is

$$(BW)_c = \sqrt{2} (BW)_p = \frac{0.225}{L} \sqrt{\omega_0^2 M^2 - \frac{R_1^2 + R_2^2}{2}} \quad (86)$$

$$= \sqrt{2} f_0 \sqrt{K^2 - \frac{1}{2} \left(\frac{1}{Q_1^2} + \frac{1}{Q_2^2} \right)} \quad (87)$$

If $R_1 = R_2 = R$,

$$(BW)_c = \frac{0.225}{L} \sqrt{\omega_0^2 M^2 - R^2} = \sqrt{2} f_0 \sqrt{K^2 - \frac{1}{Q^2}} \tag{88}$$

If Q_1 and Q_2 are both large,

$$(BW)_c = \sqrt{2} K f_0 \quad (\text{approx}) \tag{89}$$

Band width at point where gain is n db down from center gain of double peak curve when $L_1 = L_2 = L$ is $(BW)_d$ in Fig. 50:

$$(BW)_d = \frac{0.1125 \sqrt{2\omega_0^2 M^2 - R_1^2 - R_2^2 + \sqrt{2\omega_0^2 M^2 - R_1^2 - R_2^2 + 4D(R_1 R_2 + \omega_0^2 M^2)}}}{L} \tag{90}$$

where D is a function of n to be read from Fig. 51. Equation (90) is approximate and is not accurate for points more than about 10 db down from center gain.

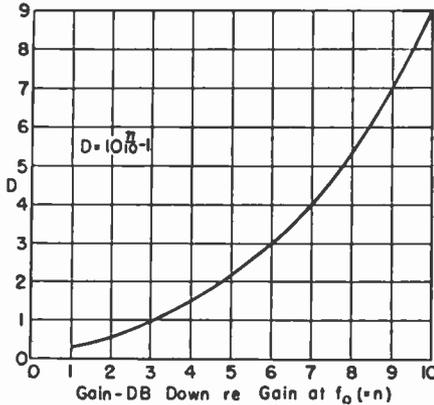


FIG. 51. Values of D vs. n for use in Eqs. (91) and (92).

If $R_1 = R_2 = R$,

$$(BW)_d = \frac{0.1592 \sqrt{\omega_0^2 M^2 - R^2 + \sqrt{\omega_0^2 M^2 - R^2 + 1D(R^2 + \omega_0^2 M^2)}}}{L} \tag{91}$$

Coupling for band width of W cycles at n db down from center gain of double peak curve when $L_1 = L_2 = L$ is

$$\omega_0 M = 2.507 \sqrt{\frac{LW}{D} \sqrt{39.48L^2 W^2 (2D + 1) + D(R_1 + R_2)^2} - \frac{39.48L^2 W^2}{D} - R_1 R_2} \tag{92}$$

where D is a function of n to be read from Fig. 51. W is identical with $(BW)_d$ in other equations. Equation (92) is approximate and is not accurate for n larger than about 10.

SOME SPECIAL APPLICATIONS OF LCR CIRCUITS

32. Band-pass R-f Circuits. If two identical tuned circuits are capacitively or inductively coupled (Fig. 52), the circuit acts as a band-pass filter with a band width

approximately

$$f_s = f_1 - f_2 = \frac{\sqrt{X_m^2 - R^2}}{2\pi L} \quad (93)$$

The band width varies with the tuning, increasing with the frequency in the inductive case, and decreasing with the frequency in the capacitive case (Fig. 52). These opposing effects may be combined in the manner shown in Fig. 52, so that the band

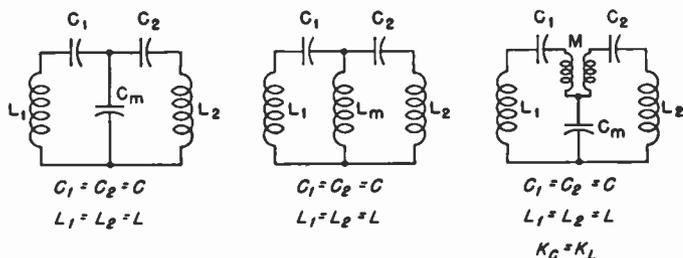


FIG. 52. Coupled circuits as band-pass filters.

width is maintained substantially constant while the circuits are tuned over a wide range of frequency by adjustment of C_1 and C_2 .

Uehling¹ has shown that this condition obtains when

$$X_{m_n} = \pm \sqrt{R_n^2 + 4\pi^2 L^2 f_s^2} \quad (94)$$

where R_n is the resistance and L the total inductance of each branch and f_s is the band width. With X_m computed for the two boundary frequencies f_a and f_b of the tuning range, the values of M and C_m required are given by

$$M = \frac{X_{m_b} f_b - X_{m_a} f_a}{2\pi(f_a^2 - f_b^2)} \quad (95)$$

$$C_m = \frac{f_a^2 - f_b^2}{2\pi f_a f_b (X_{m_a} f_b - X_{m_b} f_a)} \quad (95a)$$

Representative values of M and C_m for $f_a = 1,500$ kc, $f_b = 550$ kc, $R_n = 30$ ohms, $R_s = 10$ ohms, $L = 200 \times 10^{-8}$ henry, and $f_s = 10$ kc, which are typical constants of broadcast circuits, are

$$M = 3.2 \times 10^{-6} \text{ henry} \quad \text{and} \quad C_m = 0.06 \text{ } \mu\text{f}$$

The inductive coupling M must be negative so that its effect will be additive to that of C_m . This may be obtained by winding the coils M (Fig. 52) of two wires side by side and by connecting the "start" ends of the coils to C_1 and C_2 and the "finish" ends to C_m .

33. Decoupling Filters. When the plate current for several tubes of a high-gain amplifier is obtained from a single source, the internal resistance of the source is common to all the plate circuits and is likely to act as a coupling between stages. Similar couplings may exist through a bleeder circuit when screen voltage for two or more tubes is taken from a common tap or through a bias resistor common to the control-grid circuits of several tubes. To reduce such stray couplings to negligible amounts, decoupling filters are generally inserted in the circuits of each tube and separate bias resistors are used.

¹ *Electronics*, September, 1930, p. 279.

A typical application of decoupling filters is shown in Fig. 53, the filter elements being indicated by heavy lines. Capacitors *C* furnish low-impedance paths back to the cathodes for the signal currents flowing in the grid, screen-grid, and plate circuits, while the high-impedance resistors *R* and chokes in the leads to the voltage divider prevent any appreciable flow of signal currents in that direction. The choice of values for these resistors and chokes depends principally upon the currents in the leads and the permissible d-c voltage drop in each filter. The impedance of each by-pass capacitor should be not more than 10 per cent of that of the associated resistor or choke, at any frequency for which the amplifier is designed to operate. On the other hand, the value of *C* should not be so large in any filter that "blocking" or motorboating occurs due to too high a time constant.

The value of each cathode resistor, when separate biasing resistors are used, is equal to the bias required, divided by the total cathode d.c. of that tube. The screen-grid filter resistors serve as voltage-dropping resistors as well as filters, and their values are determined by the *IR* drops required for correct screen voltages.

34. Circuits for Obtaining Out-of-phase Voltages and Currents. Two circuits producing voltages 90 or 180 deg out of phase are shown in Figs. 54 and 55 with their vector diagrams. These are often useful in circuit designs and oscillograph measurements. To maintain these phase relations, high impedance circuits only should be connected across *e*₁ and *e*₂.

A circuit for obtaining currents 90 deg out of phase with each other is shown in Fig. 56.

To utilize these currents, nonreactive loads *R*_L and *R*_C are introduced, with values such that *R*_L*R*_C = *L*/*C* and *R*_L = *R*_C.

35. Frequency Discriminator Circuit. The frequency discriminator circuit shown in Fig. 57 is applied in a.f.c., f-m detection, frequency-drift indicators, etc. *L*₁*C*₁ and *L*₂*C*₂ are tuned to the same frequency and doubly coupled: (1) directly at *B* and (2) inductively by *M*. After rectification, a bias *E* is obtained which, between limits *C* and *D*, is proportional to the difference between the

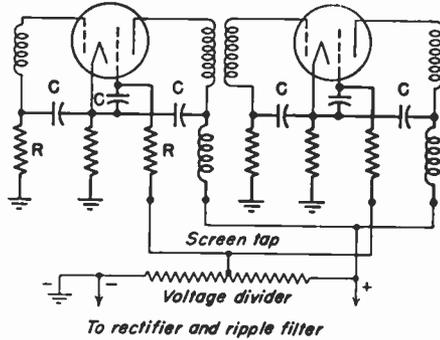


FIG. 53. Resistance-capacitance filter usage to isolate power-supply system from alternating currents.

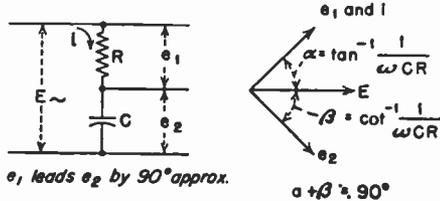


FIG. 54. *CR* circuit for obtaining out-of-phase voltages.

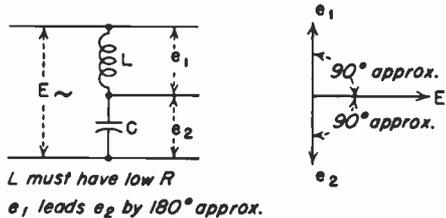


FIG. 55. *LC* circuit for obtaining out-of-phase voltages.

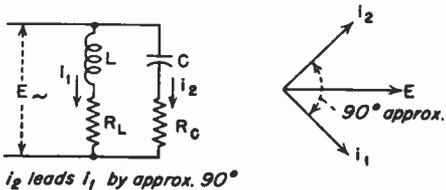


FIG. 56. Circuit for obtaining currents out of phase by 90 deg.

frequency of the input voltage and the resonance frequency of LC . The time constant of R_0C_0 should be much less than the period of one cycle of the frequency variation in the input voltage.

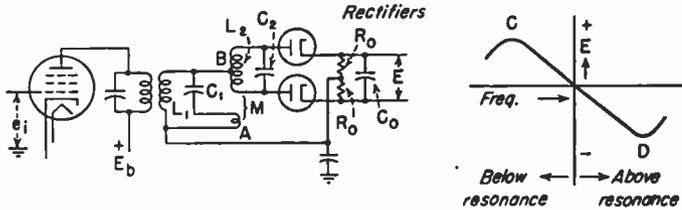


FIG. 57. Frequency discriminator and characteristic.

36. Compensation in Resistance-coupled Amplifier. In a conventional resistance-coupled amplifier (Fig. 58) the amplification falls off at low frequencies because of increasing impedance of C_c and at high frequencies because of the shunting effect of stray capacitance C_s . In wide-band amplifiers, the compensating impedances L and R_1C_1 are added. For approximately constant gain between frequency limits f_1 (low) and f_2 (high),

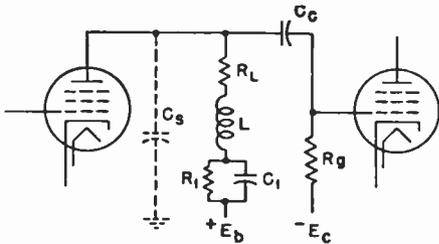


FIG. 58. Compensated resistance-coupled amplifier.

$$\left. \begin{aligned} R_L &= \frac{1}{2\pi f_2 C_s} & L &= \frac{R_L}{4\pi f_2} \\ C_1 &= \frac{1}{2\pi f_1 R_L} & R_1 &= 2R_L \end{aligned} \right\} \quad (96)$$

This type of compensation also tends to correct for phase shift near the limits f_1 and f_2 .

TRANSMISSION LINES AND NETWORKS

37. General Properties of Transmission Lines. All the relations in this section apply to any smooth transmission line with uniformly distributed constants and are exact except that radiation and end effects are not taken into account.

Characteristic impedance is equivalent to the input impedance of an infinitely long line and is

$$Z_0 = \sqrt{\frac{R + j\omega L}{G + j\omega C}} \quad \text{ohms} \quad (97)$$

or

$$|Z_0| = \sqrt{\frac{R^2 + \omega^2 L^2}{G^2 + \omega^2 C^2}} \quad \text{ohms (magnitude)} \quad (98)$$

or

$$Z_0 = \sqrt{Z_{oc} Z_{sc}} \quad \text{ohms} \quad (99)$$

where Z_{oc} and Z_{sc} are the line input impedances with the far end open- and short-circuited, respectively, and $R, L, G,$ and C are the resistance, inductance, leakage conductance, and capacitance of the line per unit of physical length. For practical purposes any convenient unit length such as 1 ft or 1 cm may be used. The same unit length must be used throughout any one computation, however.

The ratio of the currents I_1 and I_2 at points along the line one unit length apart is determined by the *propagation constant*, which is

$$P \text{ per unit length} = \log_e \frac{I_1}{I_2} = \sqrt{(R + j\omega L)(G + j\omega C)}$$

$$= A + jB \tag{100}$$

Attenuation Constant. The real part A of P determines the relative magnitude of I_1/I_2 and is

$$A = 6.141 \sqrt{\sqrt{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2)} + RG - \omega^2 LC} \quad \text{db/unit length} \tag{101}$$

Wavelength Constant. The quadrature part B of P determines the relative phase of I_1/I_2 and is

$$B = 0.707 \sqrt{\sqrt{(R^2 + \omega^2 L^2)(G^2 + \omega^2 C^2)} - RG + \omega^2 LC} \quad \text{radians/unit length} \tag{102}$$

To obtain the value in degrees per unit length, change the constant before the radical in Eq. (102) from 0.707 to 40.51.

Velocity of propagation is

$$V = \frac{\omega}{B} = \frac{2\pi f}{B} \quad \text{unit lengths/sec} \tag{103}$$

Wavelength is

$$\lambda = \frac{2\pi}{B} \quad \text{unit lengths} \tag{104}$$

Retardation time or delay is

$$t = \frac{B}{\omega} = \frac{B}{2\pi f} \quad \text{sec/unit length} \tag{105}$$

Line Terminated in Any Impedance Z_R . The input impedance is

$$Z_i = \frac{E_s}{I_s} = Z_0 \frac{Z_R \cosh Pl + Z_0 \sinh Pl}{Z_0 \cosh Pl + Z_R \sinh Pl}$$

$$= Z_0 \frac{Z_R/Z_0 + \tanh Pl}{1 + Z_R/Z_0 \tanh Pl} \quad \text{ohms} \tag{106}$$

- where l = length of line in unit lengths
- P = propagation constant per unit length
- Z_R = receiving end termination impedance
- Z_0 = characteristic impedance of line
- Z_i = input impedance of line

E_s and I_s = voltage and current at sending end

Current at any point x unit lengths from sending end in terms of sending end current is

$$I_x = I_s \left(\cosh Px - \frac{Z_i}{Z_0} \sinh Px \right) \quad \text{amp} \tag{107}$$

Voltage at any point x unit lengths from sending end in terms of sending end voltage is

$$E_x = E_s \left(\cosh Px - \frac{Z_0}{Z_i} \sinh Px \right) \quad \text{volts} \tag{108}$$

In this case, the current and voltage along the line vary from point to point because of standing waves. (See Sec. 54 for conditions in r-f lines.)

Lines Terminated in Impedance Z_0 . The input impedance for any length of line is

$$Z_i = Z_0 \quad \text{ohms} \tag{109}$$

Current or voltage at any point x unit lengths from sending end in terms of sending end current or voltage is

$$\frac{E_x}{E_s} = \frac{I_x}{I_s} = e^{-Px} = \epsilon^{-Ax} \epsilon^{-jBx} = \epsilon^{-Ax} (\cos Bx - j \sin Bx) \quad (110)$$

The ratio of voltage to current at any point x along the line is

$$\frac{E_x}{I_x} = Z_0 \quad (111)$$

In the communication field, transmission lines may be classified according to the frequency bands that they are required to transmit, as *audio-* or *radio-frequency* lines. Simplified forms of the general transmission line formulas result from the introduction of approximations that are appropriate to each case.

38. Terminating Conditions for No Reflections and Maximum Power Transfer. Conditions for maximum power transfer through a transmission line call for the source and load impedances to be the *conjugates*, respectively, of the line input impedances as seen from each end of the line with the opposite termination connected. Conditions for no reflections at the ends of the line call for source and load impedances to be *equal* to the line characteristic impedance. These conditions are mutually incompatible if the line Z_0 is not pure resistance. In the general case where Z_0 has a reactance component, the best terminating impedances are a compromise between efficient energy transfer and the avoidance of reflections. Fortunately, in practical r-f lines Z_0 is very nearly a pure resistance so that both conditions are both satisfied by making the source and load impedances pure resistances each equal to Z_0 .

Radio-frequency lines whose terminal impedances are not equal to Z_0 are inefficient for two reasons: (1) they do not deliver maximum power to the load, and (2) some energy is reflected at each mismatched junction. Any impedance irregularity along the line also produces reflections and should usually be avoided. Thus, bends in an open-wire line should be gradual, with a minimum radius of about $\lambda/4$, and the line should be kept clear by at least $\lambda/4$ from large masses of conducting or dielectric materials, including ground.

The amount of reflection at the junction of a line terminated in impedance Z_T which is not equal to Z_0 is expressed by the *coefficient of reflection*

$$k = \frac{Z_T - Z_0}{Z_T + Z_0} = \frac{(Z_T/Z_0) - 1}{(Z_T/Z_0) + 1} \quad (112)$$

In lines whose sole purpose is the conveyance of energy from one terminal to the other, reflections are objectionable because, with each passage of the wave back and forth along the line, additional energy is lost through ohmic and radiation losses. In addition, interference between the original and the reflected waves produces standing waves of voltage and current whose magnitudes determine the current-carrying capacity and insulation that must be provided in the construction of the line. This magnitude is expressed by the *standing wave ratio* (SWR),

$$\text{SWR} = \frac{E_{\max}}{E_{\min}} = \frac{I_{\max}}{I_{\min}} = \frac{1 + k}{1 - k} \quad (113)$$

where k is the coefficient of reflection defined in Eq. (112).

39. Audio-frequency Lines. In open-wire lines and large-gage cables, G is negligible, so that

$$A = 6.14 \sqrt{\omega C \sqrt{R^2 + \omega^2 L^2} - \omega^2 LC} \quad \text{db/unit length (approx)} \quad (114)$$

and

$$B = 0.707 \sqrt{\omega C \sqrt{R^2 + \omega^2 L^2} + \omega^2 LC} \quad \text{radians/unit length (approx)} \quad (115)$$

In small-gage cables, both L and G become negligibly small, and

$$A = 15.39 \sqrt{fRC} \quad \text{db/unit length (approx)} \quad (116)$$

and

$$B = 1.772 \sqrt{fRC} \quad \text{radians/unit length (approx)} \quad (117)$$

In both cases the loss is seen to vary with frequency. The transmission-loss frequency characteristics of various kinds of a-f circuits are shown in Fig. 60, and other characteristics of typical audio lines are shown in Table 2.

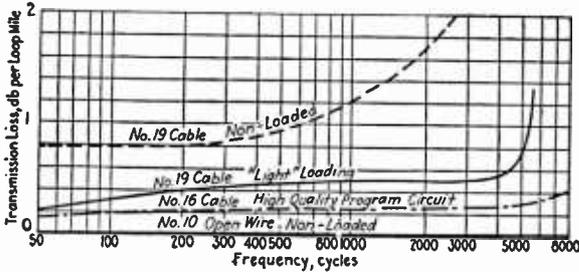


FIG. 59. Transmission-loss characteristics of various audio-frequency circuits.

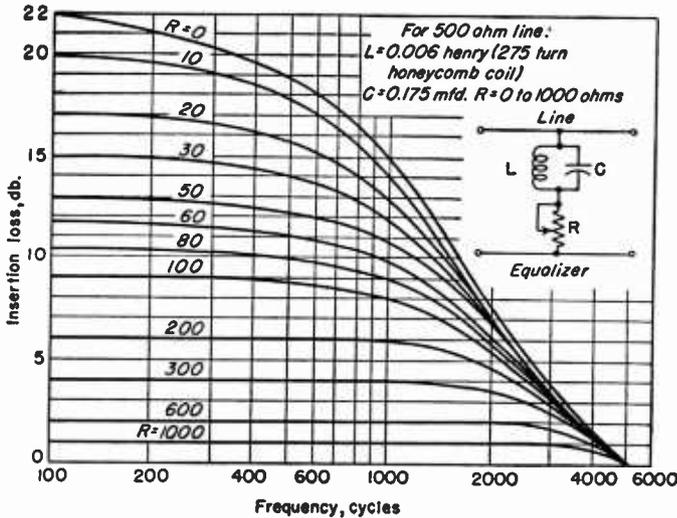


FIG. 60. Attenuation-frequency characteristic of equalizer shunted across a 500-ohm circuit.

40. Equalization of Transmission-loss Characteristic. From the curves in Fig. 59 it is evident that, if a band of frequencies is transmitted over a line, the higher frequencies will suffer more attenuation than the low frequencies, resulting in distortion. The prevention of this condition necessitates the use of attenuation equalizers in high-quality circuits. A typical 5,000-cycle equalizer for this purpose and its transmission-loss curves are illustrated in Fig. 60, and the curves for the bare line, equalizer alone, and the equalized line are shown in Fig. 61. The equalizer is usually connected in shunt across the receiving end of the line, preceding other apparatus.

Table 2. Characteristics of Some A-f Circuits
(Per mile of line at 1,000 cycles)

Type of circuit	Distributed constants per loop mile				Propagation constant		Z_0		λ , miles	Velocity, miles/sec	Loss, db/mile
	R, ohms	L, henrys	G, mhos	C, farads	A*	B, radians/mile	Magnitude, ohms	Angle, deg			
No. 8 open-wire, 18 in spacing	6.74	0.0038	0.29	0.00825	0.0050	0.0355	693	- 7.72	177	177,000	0.044
No. 16 cable, NL	42.1	0.001	1.5	0.062	0.0842	0.0974	330.6	-40.65	64.51	64,500	0.73
No. 19 cable, NL	85.8	0.001	1.5	0.062	0.1249	0.1338	470.1	-42.8	46.93	46,900	1.08
No. 16 cable, R22 load	43.1	0.040	1.5	0.062	0.0273	0.3139	809.1	- 4.76	20	20,000	0.24

* A is in nepiers per mile; multiply by 8.686 to convert to decibels per mile.

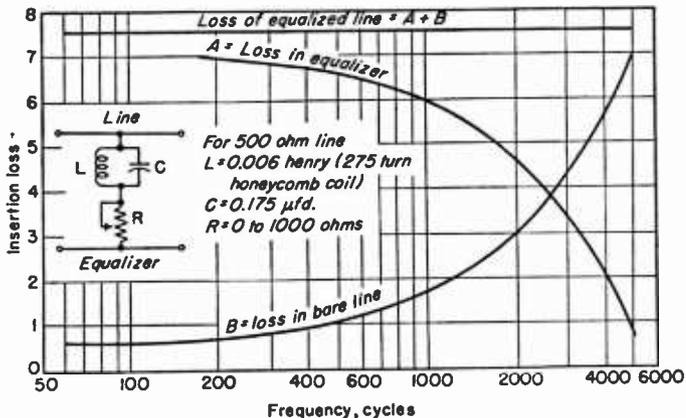


FIG. 61. Attenuation equalizer for short cable circuits.

41. Artificial Lines. An artificial line is a compact network of lumped impedances to simulate the electrical characteristics of an actual line. Such a network having approximately the characteristics of an unloaded cable or open-wire circuit may be constructed as shown in Fig. 62 and is useful in laboratory measurements and investigations.

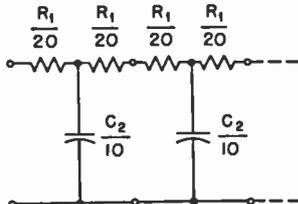


FIG. 62. Two sections of artificial nonloaded cable.

The constants R_1 and C_2 are the loop resistance and capacitance of the full length of the line to be represented. For standard cable $R_1 = 88$ ohms and $C_2 = 0.054 \mu\text{f}$ per loop mile; values for various other lines are given in Table 2. As the similarity between the artificial and the actual line increases with the number of sections in the former, it is preferable to use at least ten sections, and not more than 1 mile of cable or 10 miles of open wire should be represented by one section. The end sections should be "mid-series" terminated; *i.e.*, their series impedances should be one-half that of the internal sections.

42. Delay Networks. Three typical networks are shown in Fig. 63. It is assumed that each is fed from a source whose internal impedance is $R = \sqrt{L/C}$ and that it is terminated in a pure resistance of the same value. The delayed signal output is taken from a variable tap and fed to any device having an internal impedance which is large compared to R . Since networks I and II are respectively high- and low-pass filters of the constant-K type, they have amplitude attenuation characteristics as shown in Sec. 62. Network III has no cutoff frequency if coupling between the two halves of each coil is perfect. Relative delay vs. frequency characteristics of the three types of networks (using the same values of L and C in each case) are shown in Fig. 64. The following relations are based on resistanceless networks and hold within about 1 per cent if the coil Q is 20 or more and the capacitor Q is 100 or more.

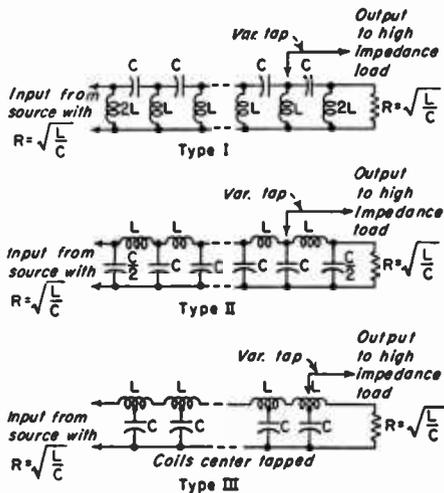


FIG. 63. Delay networks.

$$R = \sqrt{\frac{L}{C}} \quad (\text{approx}) \quad \text{for all three networks} \quad (118)$$

Network I. Amplitude attenuation is large up to

$$f_c = \frac{1}{4\pi \sqrt{LC}} \quad \text{cycles} \quad (119)$$

and is zero thereafter (high-pass).

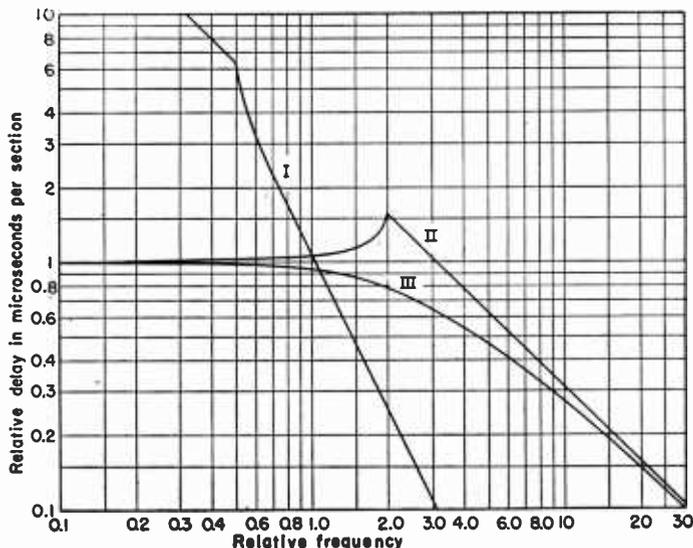


FIG. 64. Delay vs. frequency characteristics of three types of delay networks using the same LC in each case.

Delay for $f \leq f_c$ is

$$t = \frac{1}{2f} \quad \text{sec/section} \quad (120)$$

Delay for $f > f_c$ is

$$t = \frac{2}{\omega} \sin^{-1} \frac{1}{2\omega \sqrt{LC}} \quad \text{sec/section exactly (angle in radians)} \quad (121)$$

$$= \frac{1}{\omega^2 \sqrt{LC}} + \frac{1}{\omega} \left[0.04167 \left(\frac{1}{\omega \sqrt{LC}} \right)^3 + 0.00468 \left(\frac{1}{\omega \sqrt{LC}} \right)^5 + 0.000698 \left(\frac{1}{\omega \sqrt{LC}} \right)^7 \right] \quad \text{sec/section (approx)} \quad (122)$$

Delay for $f \gg f_c$ is

$$t = \frac{1}{\omega^2 \sqrt{LC}} \quad \text{sec/section (approx)} \quad (123)$$

Network II. Amplitude attenuation is zero up to cutoff frequency

$$f_c = \frac{1}{\pi \sqrt{LC}} \quad \text{cycles} \quad (124)$$

and is large thereafter (low-pass).

Delay for $f < f_c$ is

$$t = \frac{2}{\omega} \sin^{-1} \frac{\omega}{2} \sqrt{LC} \quad \text{sec/section exactly (angle in radians)} \quad (125)$$

$$= \sqrt{LC} + \frac{1}{\omega} [0.04167(\omega \sqrt{LC})^3 + 0.00468(\omega \sqrt{LC})^5 + 0.000698(\omega \sqrt{LC})^7] \quad \text{sec/section (approx)} \quad (126)$$

Delay for $f \ll f_c$ is

$$t = \sqrt{LC} \quad \text{sec/section (approx)} \quad (127)$$

Delay for $f \geq f_c$ is

$$t = \frac{1}{2f} \quad \text{sec/section (approx)} \quad (128)$$

Network III. Amplitude attenuation is zero at all frequencies for 100 per cent coupling between the two halves of each coil. In practice, this condition cannot be attained, and the attenuation will increase gradually with frequency.

Delay for all frequencies is

$$t = \frac{1}{\omega} \cos^{-1} \frac{4 - \omega^2 LC}{4 + \omega^2 LC} \quad \text{sec/section (angle in radians)} \quad (129)$$

By reference to Fig. 64 it is evident that the delay for networks II and III is constant to within ± 5 per cent for all frequencies up to that indicated as relative frequency = 1. This value corresponds to $f = 1/2\pi \sqrt{LC}$ for the network characteristics shown in the figure.

The principal use of network I is to obtain large delays at a single frequency, since in the vicinity of $f = 0.55/2\pi \sqrt{LC}$ (relative frequency = 0.55 in Fig. 64) it produces about four times as much delay per section as II or III with the same coils and capacitors.

A practical application of a delay network of type II, as used in the formation of

television synchronizing pulses, is illustrated in Fig. 65. With the constants shown the delay is $0.235 \mu\text{sec}$ per section for all frequencies up to about 680 kc.

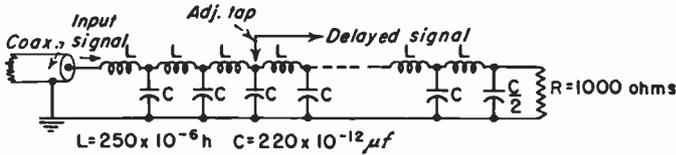


FIG. 65. Example of low-pass delay network used in television pulse circuits.

43. RC Filter for Small Currents. An economical RC filter for small currents as suggested by Scott¹ is shown in Fig. 66. A special feature is the shunting circuit through R_A to feed voltage 180 deg out of phase to the point X. This can be adjusted to give a very high attenuation at one particular frequency which it is desired to eliminate. As shown, this filter is low-pass; a similar high-pass structure can be made by transposing the R's and C's.

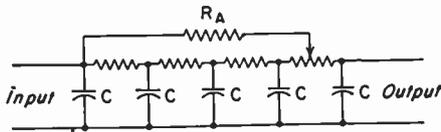
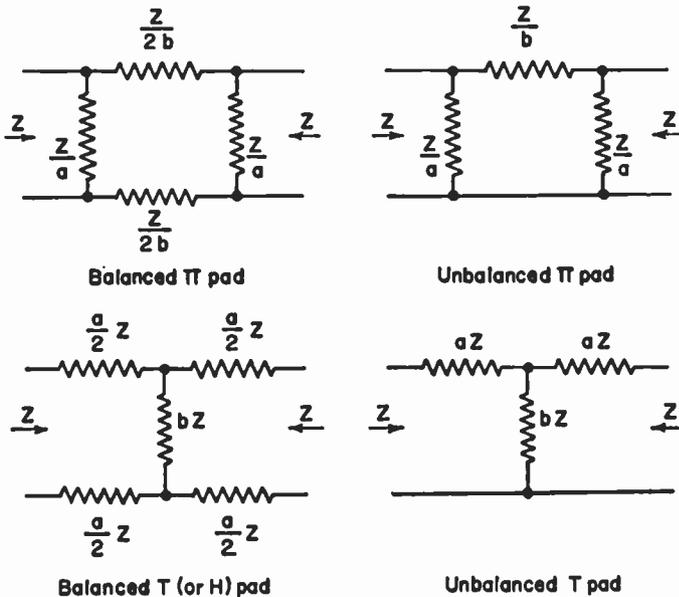


FIG. 66. RC filter for small currents.

44. Resistance Pads. Resistance pads are artificial lines whose series and shunt elements are pure resistances and are used principally as attenuators in a-f circuits.



$Z = \text{Input impedance} = \text{Output impedance}$

FIG. 67a. Pads to be used between equal impedances.

The amount of loss caused by insertion of a pad in a circuit may be accurately computed and is independent of frequency if the terminating impedances are resistances.

Either π or T structures may be used as pads, as shown in Fig. 67a. Both are elec-

¹ Scott, H. H., *Electronics*, August, 1939.

trically equivalent, but for identical values of loss and impedance one type may require resistors of more convenient values than the other. A pad to be used in a circuit that is balanced to ground should be of the balanced π or T type; otherwise the unbalanced network is satisfactory and requires several less resistors to build.

46. Pad Design. To design a pad, three constants must be known: the input and output impedances and the loss in decibels. The input and output impedances of a pad are usually made equal to those of the circuit to be connected to it. The design procedure depends upon whether these are equal or are different from each other.

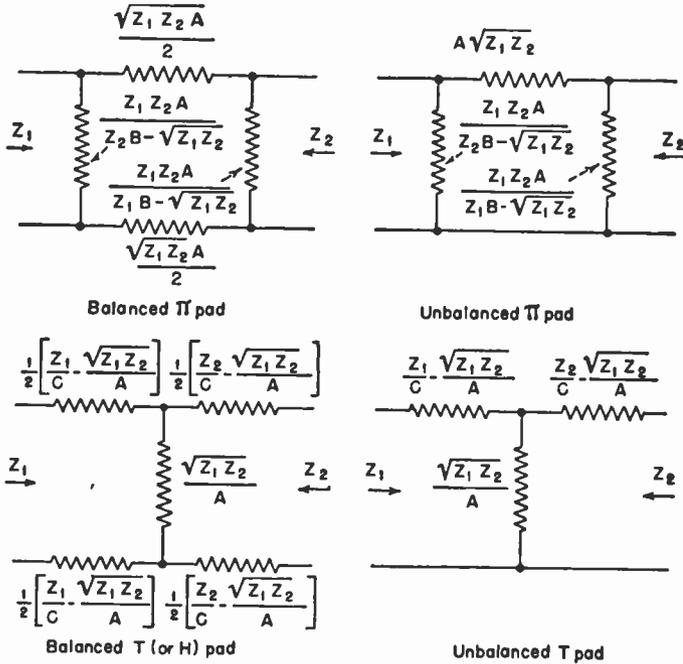


FIG. 67b. Pads to be used between unequal impedances.

1. *Equal Input and Output Impedances.* In this case, the value of each element is found by multiplying the proper constants, selected from Table 3 in connection with Fig. 67a, by the value of the input or output impedance Z in ohms.

Example: To design a 10-db, 500/500-ohm pad of the balanced T type: From Table 3, for 10-db attenuation, $a = 0.5195$ (hence $a/2 = 0.2597$) and $b = 0.7027$. Then the required resistances are $0.2597 \times 500 = 129.85$ for the series elements and $0.7027 \times 500 = 351.35$ ohms for the shunt element.

2. *Unequal Input and Output Impedances.* In this case, the design involves more computation. The value of each element is indicated by Fig. 67b, the constants of which are to be found in Table 3. The ratio of input to output impedance (or vice versa) of a pad of given loss is limited by the fact that, for large values of the impedance ratio, certain of the pad resistors would have to be negative in value if the loss of the pad were to be below a certain minimum value. The maximum impedance ratio which a 10-db pad can have, for example, is 3.018. Stated in another way, this means that, if the impedance ratio of a pad is to be 3.018, its loss must be at least 10 db. The maximum impedance ratios for various values of pad losses are also given in Table 3. These are the same for both π and T pads.

Table 3. Constants for Pads of Fig. 67

Loss, db	A	B	C	a	b	1/b	1/a	1/2b	Max ratio Z ₁ /Z ₂ or Z ₂ /Z ₁
1	0.1154	1.007	0.1150	0.0575	8.664	0.1154	17.39	0.0577	1.014
2	0.2323	1.027	0.2263	0.1146	4.305	0.2323	8.724	0.1161	1.055
3	0.3523	1.060	0.3325	0.1710	2.838	0.3532	5.848	0.1761	1.124
4	0.4770	1.108	0.4305	0.2263	2.097	0.4770	4.419	0.2385	1.228
5	0.6084	1.170	0.5192	0.2801	1.645	0.6084	3.570	0.3042	1.369
6	0.7472	1.248	0.5986	0.3323	1.339	0.7472	3.009	0.3736	1.557
7	0.8960	1.343	0.6673	0.3825	1.116	0.8960	2.615	0.4480	1.804
8	1.0570	1.455	0.7264	0.4305	0.9462	1.0570	2.323	0.5285	2.117
9	1.2320	1.586	0.7763	0.4762	0.8118	1.2320	2.100	0.6160	2.515
10	1.4218	1.738	0.8181	0.5195	0.7027	1.4218	1.925	0.7109	3.018
11	1.6324	1.914	0.8527	0.5601	0.6127	1.6324	1.785	0.8162	3.663
12	1.8659	2.117	0.8814	0.5986	0.5359	1.8659	1.670	0.9329	4.482
13	2.1223	2.346	0.9046	0.6343	0.4712	2.1223	1.576	1.0611	5.504
14	2.4067	2.605	0.9235	0.6672	0.4155	2.4067	1.498	1.2033	6.786
15	2.7230	2.901	0.9387	0.6981	0.3672	2.7230	1.432	1.3615	8.415
20	4.9522	5.052	0.9902	0.8182	0.2020	4.9522	1.222	2.4761	25.52
25	8.8612	8.918	0.9940	0.8932	0.1128	8.8612	1.119	4.4306	79.52
30	15.800	15.830	0.9980	0.9387	0.06331	15.800	1.065	7.900	250.5
35	28.094	28.112	0.9994	0.9649	0.03560	28.094	1.036	14.047	790.2
40	50.000	50.0094	0.9998	0.9802	0.020000	50.000	1.020	25.000	2,500
45	88.928	88.933	0.9999	0.9888	0.01124	88.928	1.011	44.464	7,909
50	158.1	158.102	1.0000	0.9937	0.006325	158.10	1.006	79.050	24,980
60	500	500	1.0000	0.9980	0.002000	500	1.002	250	
70	1,581	1,581	1.0000	0.9994	0.000632	1,581	1.001	790	
80	5,000	5,000	1.0000	0.9998	0.000200	5,000	1.000	2,500	
90	15,810	15,810	1.0000	0.9999	0.0000632	15,810	1.000	7,905	
100	50,000	50,000	1.0000	1.0000	0.0000200	50,000	1.000	25,000	

$$\begin{aligned}
 A &= \sinh \theta & \alpha &= \frac{1}{C} - \frac{1}{A} \\
 B &= \cosh \theta \\
 C &= \tanh \theta & b &= \frac{1}{A} \\
 \theta &= \frac{\text{loss in db}}{8.686} & \text{Max ratio } \frac{Z_1}{Z_2} \text{ or } \frac{Z_2}{Z_1} &= B^2
 \end{aligned}$$

Example: To design a 20-db 500/200-ohm pad of the unbalanced π type:

$$Z_1 = 500 \text{ ohms} \quad Z_2 = 200 \text{ ohms}$$

From Table 3, $A = 4.9522$ and $B = 5.0522$. Then,

$$\begin{aligned}
 \text{Input shunt element} &= \frac{Z_1 Z_2 A}{Z_2 B - \sqrt{Z_1 Z_2}} = 713 \text{ ohms} \\
 \text{Series element} &= \sqrt{Z_1 Z_2} A = 1,567 \text{ ohms} \\
 \text{Output shunt element} &= \frac{Z_1 Z_2 A}{Z_1 B - \sqrt{Z_1 Z_2}} = 430 \text{ ohms}
 \end{aligned}$$

46. General Properties of R-f Lines. Useful working formulas for r-f lines are obtained by making use of the fact that at radio frequencies ωL and ωC become very large in comparison with R and G , respectively. The results are approximations

whose accuracy depends upon whether R and G are treated as zero or as small quantities.

Symbols used in R-f line relations:

- L = inductance, henrys/unit length
- C = capacitance, farads/unit length
- R = series resistance, ohms/unit length
- G = shunt conductance, mhos/unit length

NOTE: L , C , R , and G are values for one unit length of two conductors.

V = actual velocity of propagation

V.F. = velocity factor, actual velocity as fraction or per cent of speed of light

v = speed of light, 3×10^8 meters/sec

$\omega = 2\pi \times$ frequency, cycles

f = frequency, cycles

f_{Mc} = frequency, megacycles

d_o = inner diam of outer conductor } both in same units

d_i = outer diam of inner conductor }

d = diam of conductor

s = spacing between conductors, center to center } both in same units

l = length of line in unit lengths

Unit length = arbitrary small length of line, as 1 cm or 1 ft

in. = (as subscript) dimensions in inches

1. *Dissipationless R-f Lines.* ($R = 0$ and $G = 0$; approximate conditions for air-insulated copper lines.)

Characteristic impedance is

$$Z_0 = \sqrt{\frac{L}{C}} = \frac{1}{vC} \quad \text{ohms} \quad (130)$$

Velocity of propagation is

$$V = \frac{1}{\sqrt{LC}} \quad \text{unit lengths/sec}$$

$$= v = 3 \times 10^8 \text{ meters/sec} \quad (131)$$

Propagation constant is

$$P = jB = j\omega \sqrt{LC} \quad (132)$$

Wavelength constant is

$$B = \omega \sqrt{LC} = \omega CZ_0 \quad \text{radians/unit length}$$

$$= \frac{\omega}{v} = 2.0944 \times 10^{-8} f \quad \text{radians/meter} \quad (133)$$

Wavelength (physical distance) is

$$\lambda = \frac{2\pi}{\omega \sqrt{LC}} = f \frac{1}{\sqrt{LC}} \quad \text{unit lengths}$$

$$= \frac{v}{f} = \frac{3 \times 10^8}{f} = \frac{300}{f_{Mc}} \quad \text{meters} \quad (134)$$

Retardation time or delay is

$$t = \frac{B}{\omega} = \sqrt{LC} \quad \text{sec/unit length}$$

$$= \frac{1}{v} = 0.333 \times 10^{-8} \quad \text{sec/meter} \quad (135)$$

2. *R-f Lines with Small Attenuation.* ($R \neq 0$ and $G \neq 0$; approximate conditions for lines with dielectrics other than air.)

Attenuation is

$$A = 4.343 \left(R \sqrt{\frac{C}{L}} + G \sqrt{\frac{L}{C}} \right) = 4.343 \left(\frac{R}{Z_0} + GZ_0 \right) \quad \text{db/unit length} \quad (136)$$

The other constants for lines with small attenuation are

$$Z_0 = \sqrt{\frac{L}{C}} \left[1 + j \left(\frac{G}{2\omega C} - \frac{R}{2\omega L} \right) \right] \quad \text{ohms} \quad (137)$$

$$V = \frac{\text{V.F.}}{\sqrt{LC}} \quad \text{unit lengths/sec} \\ = v \times \text{V.F.} \quad \text{meters/sec} \quad (138)$$

$$B = \frac{\omega \sqrt{LC}}{\text{V.F.}} \quad \text{radians/unit length} \\ = \frac{2.09 \times 10^{-9} f}{\text{V.F.}} \quad \text{radians/meter} \quad (139)$$

$$\lambda = \frac{\text{V.F.}}{f \sqrt{LC}} \quad \text{unit lengths} \\ = \frac{3 \times 10^8}{f} \times \text{V.F.} = \frac{300}{f_{\text{Mc}}} \times \text{V.F.} \quad \text{meters} \quad (140)$$

$$t = \frac{\sqrt{LC}}{\text{V.F.}} \quad \text{sec/unit length} \\ = \frac{0.333 \times 10^{-8}}{\text{V.F.}} \quad \text{sec/meter} \quad (141)$$

Although it is possible to calculate the velocity factor (V.F.) approximately for a given line, usually it is more practical to use the actually measured value. See Sec. 53 and Table 4.

The foregoing relations are equally applicable to coaxial and parallel wire lines. For lines of specified configurations, L , C , and R can be evaluated in terms of physical dimensions to give the relations that follow. G is not readily computed and, if it is not negligible, should preferably be measured.

3. *Parallel Two-wire Open R-f Line.* (Neither side grounded; copper conductors; R and G small.)

$$Z_0 = 277 \log_{10} \frac{2s}{d} \quad \text{ohms} \quad (142)$$

This holds if $s \geq 10d$ and the height above ground is $10s$ or more. Values of Z_0 for various conductor sizes and spacings are shown in Fig. 68.

$$A = 0.867 \frac{\sqrt{f_{\text{Mc}}}}{d_{\text{in.}} \times Z_0} \quad \text{db/100 ft} \quad (143)$$

where $d_{\text{in.}}$ is the conductor diameter in inches. Equation (143) is based solely on copper losses and neglects all other losses. For accurate results, measured values are preferable to calculated values.

Storage factor is

$$Q = \frac{\omega L}{R} = 883.9 d_{\text{in.}} \times \sqrt{f_{\text{Mc}}} \times \log_{10} \frac{2s}{d} \quad (144)$$

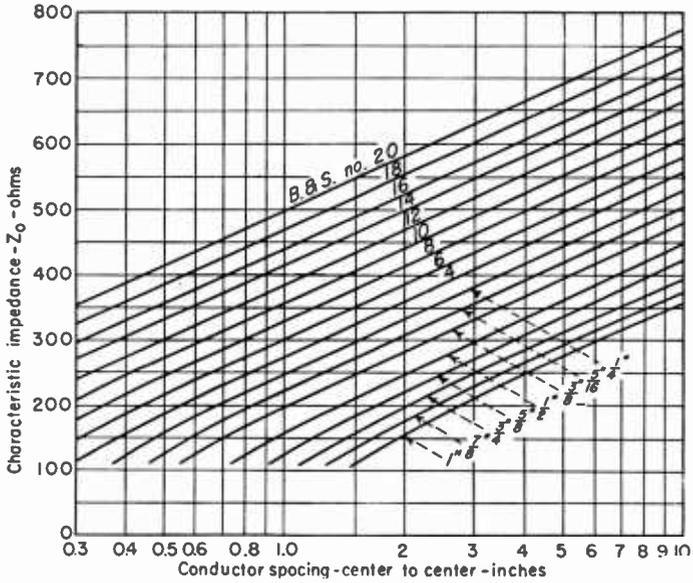


FIG. 68. Z_0 for balanced parallel copper-conductor r-f lines.

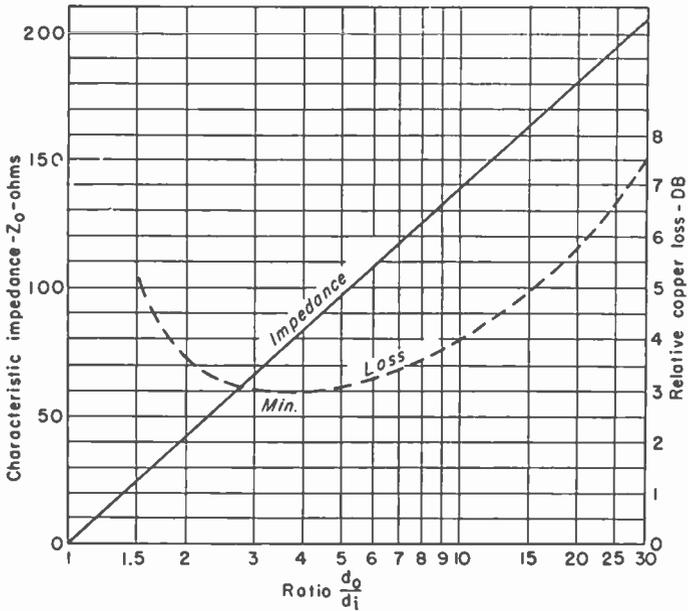


FIG. 69. Z_0 and relative copper loss vs. ratio of conductor diameters for coaxial lines.

Maximum voltage gradient is the same for both wires, occurs at the surface of each wire and at the point nearest the opposite wire, and is

$$\text{grad}_{\text{max}} = \frac{0.342E}{d_{\text{in.}} \times \log_{10} (2s/d)} \quad \text{volts/cm} \quad (145)$$

where E is the voltage between the wires at the point in question.

4. Coaxial Two-conductor R-f Line. (Copper conductors; R and G small.)

$$Z_0 = 138.5 \log_{10} \frac{d_o}{d_i} \quad \text{ohms} \quad (146)$$

Values of Z_0 vs. the ratio d_o/d_i are shown in Fig. 69.

$$A = \frac{0.00313 \sqrt{f_{M_0}} [1 + (d_o/d_i)]}{d_{o \text{ in.}} \log_{10} (d_o/d_i)} \quad \text{db/100 ft} \quad (147)$$

where $d_{o \text{ in.}}$ is the diameter of the outer conductor in inches. Equation (147) is based solely on copper losses and neglects leakage conductance G . When G cannot be neglected, measurements are preferable to calculations.

$$Q = \frac{0.8863 d_{o \text{ in.}} \sqrt{f_{M_0}} \log_{10} (d_o/d_i)}{1 + d_o/d_i} \quad (148)$$

Optimum Diameter Ratios for Copper Coaxial Lines*

Optimum quantity	d_o/d_i	Z_0 , ohms
Max antiresonant Z_0	9.185	132.9
Max Q and min attenuation (see Fig. 69).....	3.592	76.64
Max breakdown volts.....	2.718	59.93
Min temp. rise, inner conductor.....	1.835	36.38
Max power carrying capacity.....	1.648	29.94

* Smith, P. H., *Electronics*, February, 1950.

Maximum voltage gradient occurs at the surface of the inner conductor and is

$$\text{grad}_{\text{max}} = \frac{0.342E}{d_{i \text{ in.}} \log_{10} (d_o/d_i)} \quad \text{volts/cm} \quad (149)$$

where E is the voltage between inner and outer conductors and $d_{i \text{ in.}}$ is the diameter of inner conductor in inches.

5. Resistance of Copper R-f Lines. For two-wire open lines,

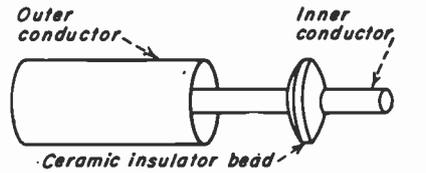
$$R = \frac{0.1996 \sqrt{f_{M_0}}}{d_{\text{in.}}} \quad \text{ohms/100 ft} \quad (150)$$

For coaxial lines,

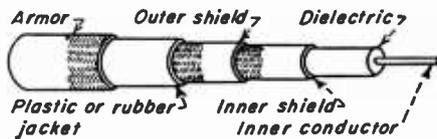
$$R = \frac{0.0998 \sqrt{f_{M_0}} [1 + (d_o/d_i)]}{d_{o \text{ in.}}} \quad \text{ohms/100 ft} \quad (151)$$

In both cases, the resistance is in terms of 100 ft of line and takes into account the fact that there are two conductors.

47. RCA Six-wire Line. A type of open-wire line frequently used for feeding a broadcast transmitting antenna which operates with one side grounded is shown in Fig. 71. The four outer wires are connected together and grounded to act as one conductor, and the two inner wires are connected together to act as the second conductor. The outer wires function as a shield and limit radiation from the line to about



(a) Air dielectric bead-insulated line



(b) Solid dielectric cable

FIG. 70. Typical construction of commercial coaxial types of r-f transmission lines.

Table 4. Characteristics of Commercial R-f Transmission Lines
(Manufacturer's ratings)

Manufacturer	Type No.		Type line	Outside dimensions, in.	Insulation	Conductor size (inner, if coaxial)	Velocity*	Z ₀	Max rms volts or power	Loss in decibels per 100 ft							
	Mfr.	RG/-								1 Mc	10 Mc	50 Mc	100 Mc	200 Mc	1,000 Mc	3,000 Mc	
Amphenol.....	21-004	8	Coaxial	0.405 diam	Polyethylene	7—No. 21	65.9	52	4,000 volts	0.130	0.520	2.10	8.50	18.0	
	21-007	11	Coaxial	0.405 diam	Polyethylene	7—No. 26	65.9	75	4,000 volts	0.115	0.470	1.90	7.70	16.0	
	21-013	17	Coaxial	0.870 diam	Polyethylene	0.188 in.	65.9	52	11,000 volts	0.038	0.180	0.85	4.00	9.6	
	21-015	19	Coaxial	1.12 diam	Polyethylene	0.250 in.	65.9	52	14,000 volts	0.030	0.142	0.70	3.40	8.3	
	21-017	21	Coaxial	0.332 diam	Polyethylene	16 Nichrome	65.0	53	2,700 volts	1.3	4.3	14.0	46.0	82.0	
	21-019	34	Coaxial	0.625 diam	Polyethylene	7—No. 21	65.9	72	5,200 volts	0.115	0.470	1.90	7.70	16.0	
	21-024	58	Coaxial	0.195 diam	Polyethylene	No. 20	65.9	53.5	1,900 volts	0.240	1.00	4.10	17.0	34.0	
	21-025	59	Coaxial	0.242 diam	Polyethylene	No. 22	65.9	73	2,300 volts	0.260	1.00	3.75	14.3	28.5	
	21-033	22	Parallel shielded	0.405 diam	Polyethylene	7—0.0152 in. each	65.9	95	1,000 volts	0.220	0.860	3.40	
	6 wire	..	Open wire	15 square†	Ceramic post	No. 6 stranded	230	500 kw	0.04	
Andrew.....	83	..	Coaxial	0.375 diam	Spaced beads	No. 12	86.0	70	250 watts‡	0.100	0.324	0.754	1.09	1.61	
	87	..	Coaxial	1.625 diam	Spaced beads	0.50 in.	97.8	66	5 kw‡	0.020	0.065	0.150	0.219	0.321	
	737	..	Coaxial	0.875 diam	Spaced beads	0.25 in.	92.0	64	2 kw‡	0.037	0.119	0.281	0.414	0.617	
	450	..	Coaxial	0.875 diam	Spaced beads	0.312 in.	93.3	51.5	3 kw‡	0.041	0.130	0.318	0.460	0.673	
	451	..	Coaxial	1.625 diam	Spaced beads	0.625 in.	95.3	51.5	10 kw‡	0.021	0.068	0.161	0.234	0.346	
	452	..	Coaxial	3.125 diam	Spaced beads	1.2 in.	92.8	51.5	42 kw‡	0.011	0.035	0.096	0.145	0.230	
Amphenol.....	453	..	Coaxial	6.125 diam	Spaced beads	2.5 in.	99.3	51.5	166 kw‡	0.0053	0.017	0.039	0.056	0.081	
	14-056	..	Parallel conductor	0.4 × 0.062	Polyethylene	7—No. 28	82.0	300	Receiver type	0.390	1.30	2.10	3.60	
	14-079	..	Parallel conductor	0.185 × 0.06	Polyethylene	7—No. 28	77.0	150	Receiver type	0.460	1.60	2.70	4.70	
	14-080	..	Parallel conductor	0.133 × 0.06	Polyethylene	7—No. 28	68.0	75	Receiver type	0.850	3.00	5.00	8.30	

* Velocity in per cent free space velocity.

† Not including mounting brackets.

‡ For line operated with SWR = 1.

Data obtained through courtesy of American Phenolic Corp., Andrew Co., and RCA.

0.015 per cent in one wavelength of line as compared to about 1 per cent for an unbalanced two-wire line of corresponding dimensions. Ground conductivity losses are also reduced by using the six-wire line. The characteristic impedance using No. 6 copper conductors is 230 ohms, and the attenuation is approximately $0.04 \sqrt{f_{Mc}}$ db per 100 ft. throughout the 550 to 1,600-ke broadcast band for a soil conductivity of 50×10^{-18} emu.

48. Impedance Relations in R-f Lines. At high frequencies the attenuation constant of a line is so small that it may be neglected in many cases to simplify impedance calculations. On this basis, Z_0 is essentially pure resistance, and the input impedance of a line terminated at its end in any impedance Z_T reduces, from Eq. (106), to

$$\begin{aligned}
 Z_i &= Z_0 \frac{Z_T \cos lB + jZ_0 \sin lB}{Z_0 \cos lB + jZ_T \sin lB} \\
 &= Z_0 \frac{Z_T \cos 2\pi l/\lambda + jZ_0 \sin 2\pi l/\lambda}{Z_0 \cos 2\pi l/\lambda + jZ_T \sin 2\pi l/\lambda} \\
 &= Z_0 \frac{Z_T/Z_0 + j \tan 2\pi l/\lambda}{1 + jZ_T/Z_0 \tan 2\pi l/\lambda}
 \end{aligned}$$

ohms (152)

in which l is the length of the line and λ is the wavelength of the signal, both measured in the same unit of length. This gives the value of $2\pi l/\lambda$ in radians. It is often convenient to substitute the electrical length of the line in degrees for $2\pi l/\lambda$, on the basis of 360 deg per wavelength. Thus a $\lambda/4$ line represents 90 electrical degrees, etc.

Actual numerical computation of input impedances by Eq. (152) is a tedious process, especially if Z_T is not pure resistance. When many such computations are to be made, the transmission line calculator of P. H. Smith¹ is a practical timesaver. This device is in the form of a circular slide rule and also yields such information as standing wave ratios and line attenuation.

Open-circuited Line. The input impedance of a no-loss line of any length open-circuited at its far end ($Z_T = \infty$) is a pure reactance

$$Z_i = -jZ_0 \cot lB = -jZ_0 \cot 2\pi l/\lambda \tag{153}$$

Short-circuited Line. The input impedance of a no-loss line of any length short-circuited at its far end ($Z_T = 0$) is a pure reactance

$$Z_i = jZ_0 \tan lB = jZ_0 \tan 2\pi l/\lambda \tag{154}$$

The reactive input impedance of a short length of line, or "stub" line, provides a handy method of balancing out unwanted reactance in the termination of a h-f line. The stub is connected in parallel across the terminating impedance and trimmed in length by trial until an increase in signal strength or similar indication shows that the desired result is obtained. Such stubs may be either open- or short-circuited at their free ends.

49. Properties of Quarter-wave R-f Lines. In dissipationless r-f lines the input impedance Z_i of a $\lambda/4$ line—and of all lines whose lengths are odd multiples of $\lambda/4$ —is the reciprocal of its terminating impedance Z_T when the two impedances are expressed

¹ Smith, P. H., Transmission Line Calculator, *Electronics*, January, 1939, and Improved Transmission Line Calculator, *Electronics*, January, 1944.

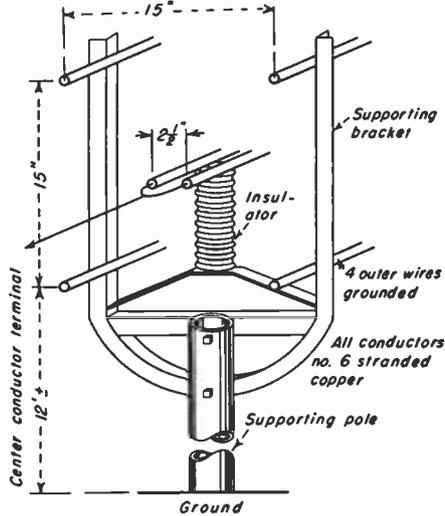


FIG. 71. Six-wire r-f line with low radiation. (Courtesy RCA.)

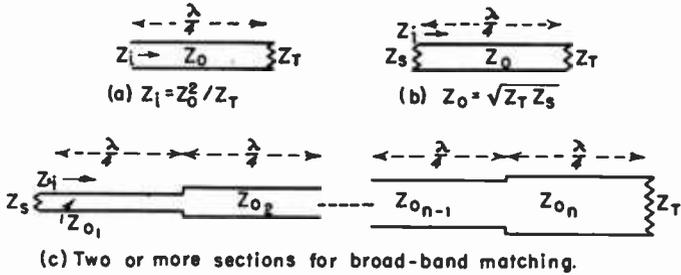


FIG. 72. Quarter-wave lines as impedance transformers. All sections are $\lambda/4$ at f_0 .
 in terms of the line characteristic impedance Z_0 (Fig. 72). Thus

$$\frac{Z_i}{Z_0} = \frac{Z_0}{Z_T} \quad \text{or} \quad Z_i = Z_0^2 / Z_T \quad (155)$$

This condition is independent of the value of the characteristic impedance Z_0 of the line.

If a $\lambda/4$ line is short-circuited at its far end ($Z_T = 0$), Z_i is very large, approaching ∞ in the case of a dissipationless line. This fact is made use of in such practical applications as the

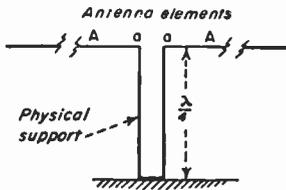


FIG. 73. Support for antenna elements A made into form of $\lambda/4$ line, effectively insulating points $a-a$ from ground.

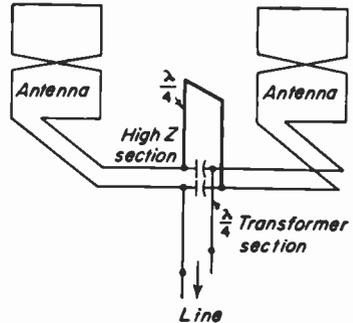


FIG. 74. Use of $\lambda/4$ short-circuited line to by-pass low-frequency currents for melting sleet without disturbing the r-f impedance of the system.

“conductive insulator,” Fig. 73, and the sleet-melting arrangement shown in Fig. 74.

If a $\lambda/4$ line is open-circuited at its far end ($Z_T = \infty$), Z_i is very small, approaching zero for a dissipationless line. A practical application is the third-harmonic suppressor shown in Fig. 75, in which an open-circuited $\lambda/4$ line A (at third-harmonic frequency) is shunted across a line carrying a signal of fundamental frequency f . The stub line effectively short-circuits the main line for currents of frequency $3f$, thereby eliminating the third harmonic. However, the effect of the stub at the fundamental frequency f must also be considered. For this frequency, the stub length is $\lambda/12$, and the line presents a shunt reactance of $-j1.732Z_0$ ohms to currents of frequency f . To balance out this reactance a second stub B , of length $\lambda/6$ and short-circuited at its far end, is also shunted across the main line at the same point. This presents an equal and opposite reactance of $+j1.732Z_0$ ohms to resonate

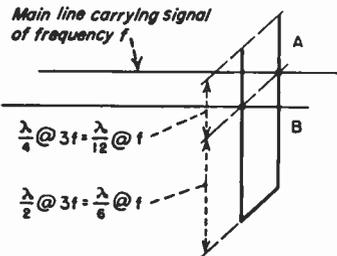


FIG. 75. Use of stub transmission lines to suppress third harmonic.

In Fig. 76 the number attached to each curve represents the total number of $\lambda/4$ sections employed. The curves shown are for a transformation ratio of 10:1. For a smaller ratio, they will be flatter; for a larger ratio, they will show larger variations within a given range of frequency. In the design of an exponentially tapered $\lambda/4$ line transformer to match an impedance Z_T to another impedance Z_o , it is convenient to use the last expression in Eq. (157) together with Table 5 which gives values of a for

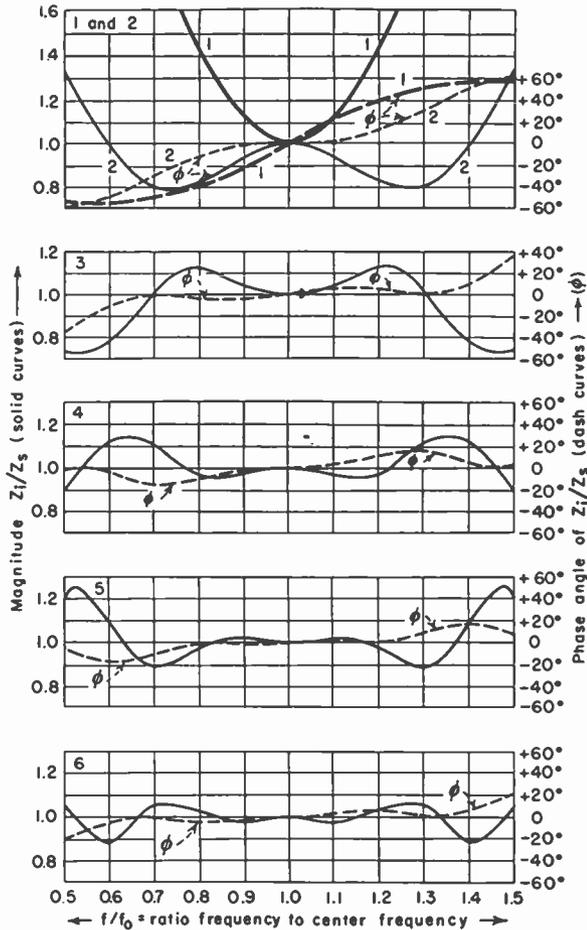


FIG. 76. Degree of wide-band impedance match obtained by use of from one to six $\lambda/4$ transformer sections between 10:1 impedance ratio.

transformers of from one to six sections. The value of f_0 should be chosen in the center of the frequency band to be covered by the system, and the physical length of each section of line should correspond to an electrical length of $\lambda/4$ at the frequency f_0 .

52. Properties of Half-wave R-f Lines. Dissipationless half-wave lines and lines whose lengths are integral multiples of $\lambda/2$ have input impedance Z_i exactly equal to the terminating impedance Z_T , independent of the Z_o of the line.

$$Z_i = Z_T \quad \text{ohms} \tag{158}$$

53. Dimensional Data for R-f Stub Lines. The range of frequencies in which the use of stub or transformer lines is practical is limited at the lower extreme by the physical dimensions of the line structure required at the longer wavelengths, and at the upper extreme by the relative inefficiency of conductor type lines as compared to wave guides.

To determine the physical length of a stub line, it is necessary to take into account the actual velocity of propagation along the line. The following relations are useful in determining physical lengths of lines for this purpose:

$$\begin{aligned} \lambda/4 \text{ length} &= \frac{75}{f_{Mc}} \times \text{V.F. meters} = \frac{246.1}{f_{Mc}} \times \text{V.F. ft} = \frac{2,952.8}{f_{Mc}} \times \text{V.F. in.} \\ \lambda/2 \text{ length} &= \frac{150}{f_{Mc}} \times \text{V.F. meters} = \frac{492.1}{f_{Mc}} \times \text{V.F. ft} = \frac{5,905.5}{f_{Mc}} \times \text{V.F. in.} \\ \lambda \text{ length} &= \frac{300}{f_{Mc}} \times \text{V.F. meters} = \frac{984.3}{f_{Mc}} \times \text{V.F. ft} = \frac{11,811}{f_{Mc}} \times \text{V.F. in.} \end{aligned}$$

where f_{Mc} = frequency, Mc

λ = 1 electrical wavelength

V.F. = velocity factor expressed as a fraction of the speed of light.

The value of V.F. must be determined for the line to be used. Representative values for common types of lines are

V.F. = 0.92 to 0.99 for open-wire air dielectric lines

V.F. = 0.79 to 0.99 for beaded coaxial lines with air dielectric

V.F. = 0.6 to 0.82 for solid dielectric lines

54. Voltage and Power Ratings of R-f Lines. In selecting a transmission line for a transmitter application, consideration must be given to voltage flashover and power limitations. Maximum voltage is limited by conductor spacing, insulation and the corona formation point, and maximum power by line losses and conductor current-carrying capacity. Voltage, current, and power are interrelated in the case of a line operated with matched terminal impedances so that standing waves do not exist:

$$P = \frac{E^2}{Z_0} = I^2 Z_0 \quad \text{watts} \tag{159}$$

where P = power carried by line, watts

E = rms voltage between conductors, volts

I = rms current in each conductor, amp

Z_0 = characteristic impedance of line, ohms

The corresponding relations for a line with standing waves are

$$E_{\max} = \sqrt{P \times Z_0 \times \text{SWR}} \tag{160}$$

$$I_{\max} = \sqrt{\frac{P \times \text{SWR}}{Z_0}} \tag{161}$$

where SWR is the standing wave ratio as defined in Eq. (113).

Corona occurs when the voltage gradient at some point in the line exceeds the breakdown potential of air (or gas in the case of gas-filled lines). Corona is not necessarily accompanied by flashover but is objectionable because the energy dissipated represents an increase in the line losses. Chemical by-products of sustained corona discharges may also tend to deteriorate the line insulation. The critical surface gradient at which corona begins to form in air at 25°C and atmospheric pressure depends upon the conductor size. For conductors of No. 0 B&S and larger, it is about 40,000 volts per cm peak, or 28,000 volts per cm rms. For smaller conductors, it is approximately $(40,000 + N \times 1,670)$ volts per cm peak, or $(28,000 + N \times 1,180)$ volts per cm rms, where N is the B&S gage size of the wire. Maximum gradients in r-f lines occur at the surfaces of the conductors in open-wire lines, and at the surface of

the inner conductor in coaxial lines. They may be evaluated by Eqs. (145) and (149). All such computations must be considered as approximations, but they are sufficiently accurate for engineering purposes if adequate safety factors are allowed.

WAVE FILTERS

55. Wave filters are forms of artificial lines, such as those of Fig. 77, purposely designed to transmit efficiently current in a desired band of frequencies and more or less completely to suppress all other frequencies. The boundary frequencies between transmission bands and attenuation bands are called *cutoff frequencies*.

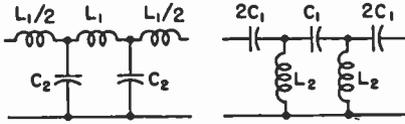


Fig. 77.

The following brief discussion of wave-filter design is intended to serve as a guide to the design of simple filters for use where the requirements are not very severe. For complete information concerning the design of filters to meet more exacting

specifications, the references listed in the bibliography at the end of this section should be consulted.

Filters are divided into four classes, according to the frequency bands which they are intended to transmit: *low pass*, *high pass*, *band pass*, and *band elimination*.

56. Losses in Filters, and Effects of Dissipation. The elements of ideal wave filters are always pure reactances; practically, however, some dissipation must always be tolerated owing to the resistance of coils and condensers, but this is made as small as possible by employing high-*Q* elements.

The terminating impedances of a filter are usually resistances equal in value to the image impedances of the filter. Then the loss within the transmitted bands (except near the cutoff frequency) is mainly due to dissipation in the elements and is usually small. In the vicinity of cutoff and the point of maximum attenuation, the total insertion loss of a filter involves the reflection and interaction losses as well as the attenuation. The loss elsewhere in the attenuated bands is very nearly the sum of the attenuation constants of the various sections, minus a gain of approximately 6 db which is due to reflections resulting from impedance mismatches occurring in these regions. Methods for the exact calculation of filter losses are beyond the scope of this handbook but are available in the published works of Zobel, Johnson, and Shea.

57. The Basic Filter Section. The basis of filter design is the full *L* section, consisting of a series element Z_1 and a shunt element Z_2 as shown at *a* in Fig. 78. The relation of such a section to an infinite line is also indicated. In a wave filter, where the number of sections is finite and small instead of infinite, symmetrical sections are used. These are either *T* or π networks as shown at *b* and *c* in Fig. 78. The *T* section may be considered as being cut from the infinite line (Fig. 78*a*) at the mid-points (*a-a*) of two consecutive series elements Z_1 and is said to be "mid-series terminated." The π section may be considered as being cut at the midpoints (*b-b*) of two consecutive shunt elements and is said to be "midshunt terminated." (To form a mid-shunt termination, each full-shunt element is replaced by an equivalent two impedances in parallel, each of value $2Z_2$.) Either a *T* or π section may be divided into pairs of equivalent half sections as shown at *d* and *e* in Fig. 78.

58. Types of Sections. 1. *Constant-K Sections.* The simplest and most common type of filter section is that in which the impedances Z_1 and Z_2 are so related that their product is a constant

$$Z_1 \times Z_2 = K^2$$

at all frequencies. From this it derives its name "constant-*K*" section. The configuration and circuit constants of the four classes of constant-*K* sections are shown in the filter-design formulas in Sec. 62. The image impedances of mid-series and mid-shunt terminated constant-*K* sections within the transmission bands are functions of frequency, but each approaches the value *K* at some frequency within the band.

The value K is therefore taken as the nominal resistance of the constant- K section for design purposes. If a constant- K section is used with one or both of its terminals connected to a pure resistance of value $R = K$, the impedances will be mismatched for all frequencies within the transmitted band except one, and the actual insertion or transmission loss of the filter will be increased by reflection losses at the terminations. This causes an even more gradual cutoff for the constant- K section than its attenuation curve would indicate.

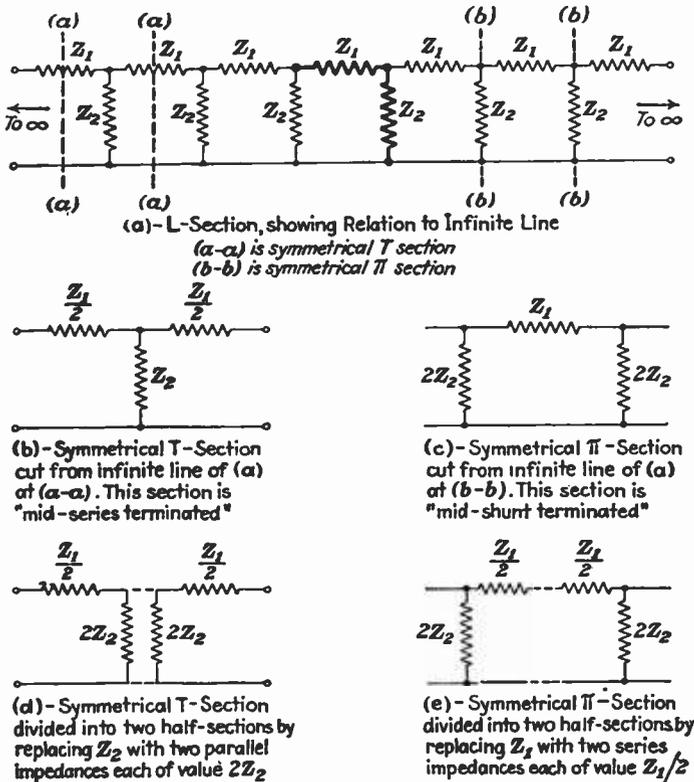


FIG. 78. Equivalence of T and π networks.

2. *m-Derived Sections.* In many filters, a sharper cutoff than that given by a constant- K type of structure is required. Such a characteristic may be realized in the so-called *m-derived section*, which is due to Otto J. Zobel.¹ This type of section is derived from the constant- K section as a prototype but is made to have sharper cutoff than the prototype by the addition of impedance elements in either the shunt or series arms so that infinite attenuation occurs at some frequency beyond cutoff. Each impedance of the *m*-type section is related to those of the constant- K section by a factor which is a function of a constant m . The latter is in turn a function of the ratio between the frequency of infinite attenuation and the cutoff frequency and may have any value between 0 and plus 1. The sharpness of cutoff increases as m approaches 0. This effect is illustrated in Fig. 79 for various values of m . It will be noted that, when m is equal to 1, the structure is identical with the constant- K

¹ Bell System Tech. J., January, 1923.

structure. Also, from Fig. 79, it appears that from the viewpoint of obtaining a uniform degree of attenuation throughout the attenuated band the combination of a constant- K section ($m = 1$) (having gradual cutoff but large attenuation remote from cutoff) with one having a small value of m and sharp cutoff ($m = 0.3$, for example) would be desirable. This principle is valuable in the design of composite filters.

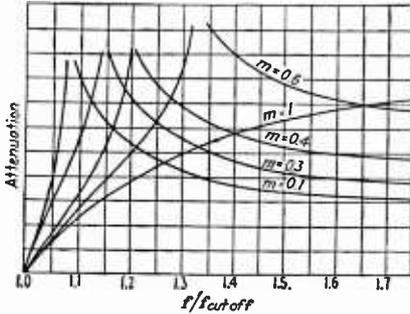


FIG. 79. Effect of m upon sharpness of cutoff in a low-pass filter structure.

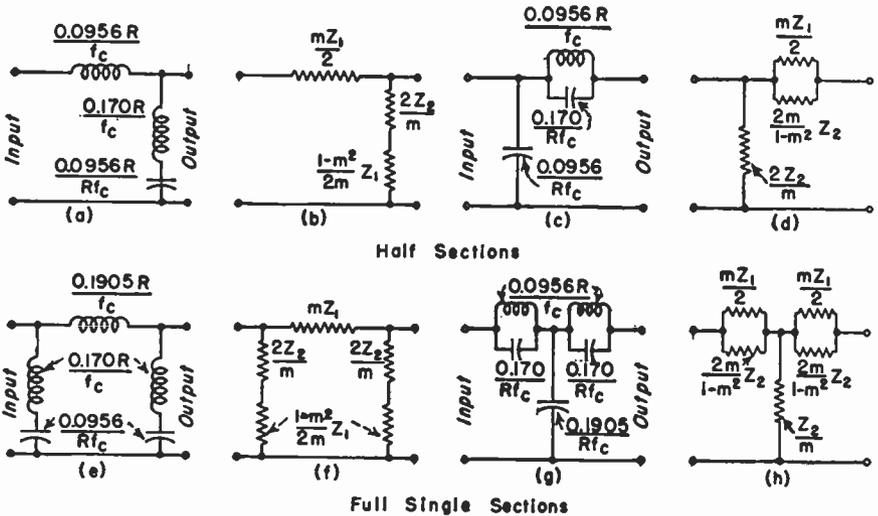
attenuation in the rejected band depends upon the number of filter sections used, and the shape of the transmission curve depends upon the types of sections employed.

60. One-half- and One-section Filters. If a half section or one full section is used alone as a filter and the requirements regarding the cutoff are not too sharp, an

3. *Shunt-derived and Series-derived m Sections.* Two forms of m -derived sections exist; if the extra impedance is added to the shunt arm, the section is called *series derived*, while, if it is added to the series arm, the section is called *shunt derived*. (See illustrations of derived sections under Filter-design Formulas, Sec. 62.)

59. Assembly of Sections into Filters.

A filter may consist of any number of sections from a single one-half section to five or six full sections, depending on the amount of attenuation of unwanted frequencies required. The amount of at-



Series-Derived m -Type Section and Half-Section, Mid-Shunt terminated.	f_c = Cut off Frequency R = Terminating Resistance $m = 0.6$ in Examples	Shunt-Derived m -Type Section and Half-Section, Mid-Series terminated.
--	--	--

FIG. 80.

m -derived section is usually preferable, with $m = 0.6$. This will provide the best impedance match with resistance terminations. Either of the structures shown in Fig. 80 is suitable for use with terminations of resistance R .

61. Multisection Filters. Filters having more than one section are of two types:

A *uniform* filter is one in which all sections are identical with the exception of the end sections. The latter are ordinarily half sections suitable for connecting the filter to its terminating resistances.

A *composite* filter is one made up of two or more sections having different characteristics, each of which is designed to contribute some especial property to the characteristic of the filter as a whole. For example, one section which has sharp cutoff but a diminishing attenuation beyond cutoff may be combined with another section having a gradual cutoff and increasing attenuation beyond as shown at I and II in Fig. 81. The resulting composite structure will then have both sharp cutoff and high attenuation beyond, as shown at III. In general, constant-*K* sections have gradual cutoffs with increasing attenuation beyond, while *m*-sections with small values of *m* have the sharpest cutoff characteristics. Still other types of sections may be added to match impedances at the junctions of the filter and its terminating resistances, or to further alter the transmission characteristics.

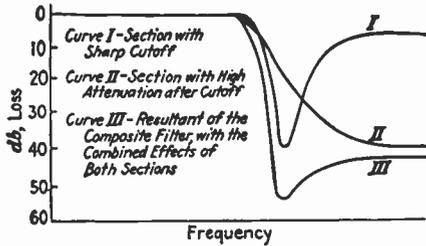


FIG. 81. Transmission curves for composite low-pass filter.

In a composite filter it is essential that the image impedances be matched at each junction of the component sections, to avoid reflection losses which would impair the transmission curve of the filter. Likewise, the end terminations of the filter should be as nearly as possible match the terminating resistances. One of the principal advantages of the *m*-type structure is that its image impedances can be made identical with other *m*-type sections or with constant-*K* sections; or they can be made to approximate resistances over the transmission band for terminating purposes. A complete analysis of the impedance conditions within a wave filter is not possible in the limited space available here but may be found in the references listed at the end of this chapter. The following will suffice as working rules in designing simple filters for ordinary requirements:

End Terminations. Resistance. A mid-shunt termination of a series-derived *m*-type section or half section, or a mid-series termination of a shunt-derived section or half section, with *m* = 0.6 in either case.

For Parallel or Series Connection with Other Filters. An 0.8 series constant-*K* section or half section (i.e., one terminated in a series arm equal to 0.8 of a full series arm, *Z*₁).

Here, as well as in the two preceding paragraphs, the image impedance of the internal section next to the end section in either case must match the image impedance at the inner terminals of the latter, in accordance with the following.

Internal Junctions. The following terminations of the types of filter sections for which formulas are given in Sec. 62 may be joined together without impedance mismatches at the junction points:

Mid-series termination of constant-*K* type to mid-series termination of series-derived *m* type.

Mid-shunt termination of constant-*K* type to mid-shunt termination of shunt-derived *m* type.

Mid-series termination of constant-*K*, series-derived *m* type or shunt-derived *m* type, to mid-series termination of another section of the same type.

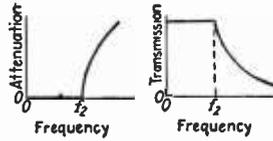
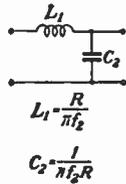
Mid-shunt termination of constant-*K*, series-derived *m* type or shunt-derived *m* type, to mid-shunt termination of another section of the same type.

NOTE: In the latter two cases, the values of *m* in the two sections to be joined, if they are of the *m* type, may be, and frequently are, different. Both sections must be of the same type and termination, however.

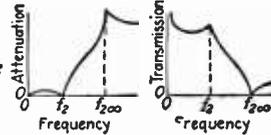
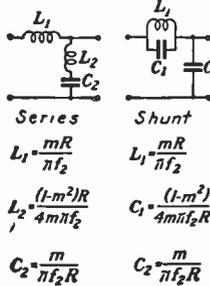
62. Filter-design Formulas. Formulas for calculating the capacitances and inductances of constant-*K*, series-derived *m*-type and shunt-derived *m*-type basic sections

I. LOW PASS FILTERS

(a)-Constant *K* Type



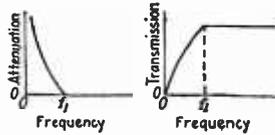
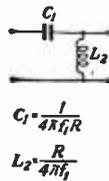
(b) *m*-Derived Type



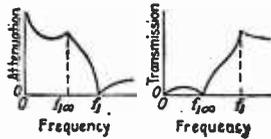
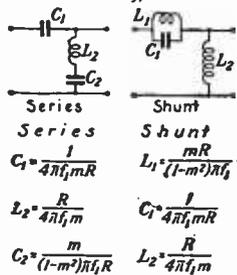
$$m = \sqrt{1 - \frac{f_2^2}{f_{200}^2}}$$

II - HIGH PASS FILTERS

(a)-Constant *K* Type



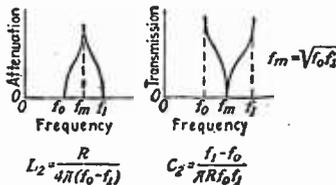
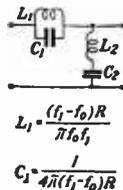
(b)-*m*-Derived Types



$$m = \sqrt{1 - \frac{f_{100}^2}{f_1^2}}$$

III - BAND ELIMINATION FILTERS

(a)-Constant *K* Type

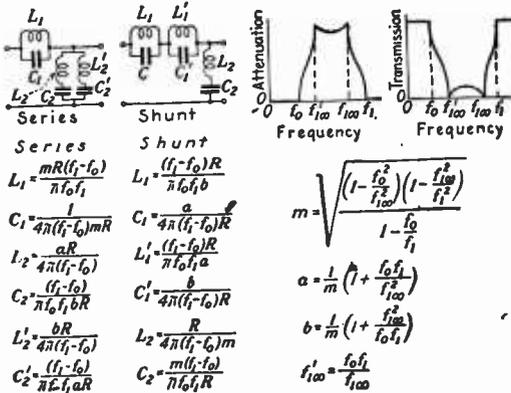


$$L_2 = \frac{R}{4\pi (f_1 - f_0)}$$

$$C_2 = \frac{f_1 - f_0}{\pi R f_0 f_1}$$

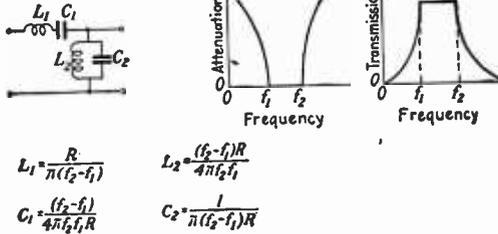
III - BAND ELIMINATION FILTERS (continued)

(b) - *m*-Derived Types

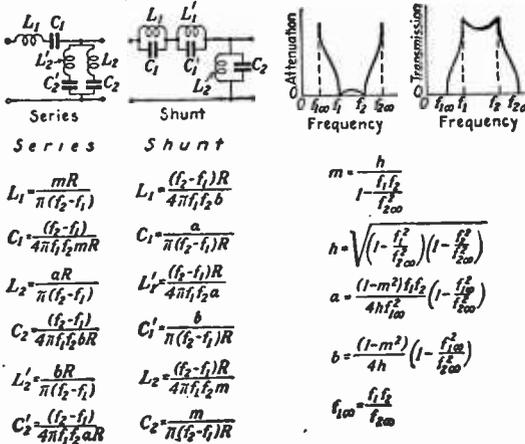


IV. BAND PASS FILTERS

(a) - Constant *K* Type



(b) - *m*-Derived Types



are given on pages 250 and 251. These are expressed in terms of R , the terminating resistances, the factor m , and the values of f_c , the cutoff frequency, and other critical frequencies. These factors must be predetermined on the basis of the filter requirements and the considerations outlined above.

Examples of Filter Design: 1. Single-section Filter. Required: High-pass single-section filter to be connected between resistance terminations of $R = 1,000$ ohms, with a cutoff frequency of 1,000 cycles and maximum attenuation occurring at 800 cycles.

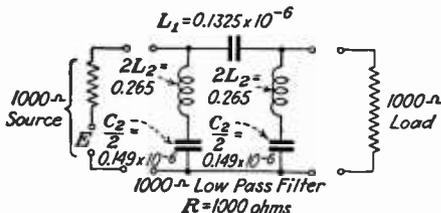


FIG. 82. Example of single-section filter.

To secure the attenuation peak at 800 cycles, an m -type filter section is required. Either the shunt- or series-derived type may be used. Choosing the latter, we have from the filter formulas II (b), Sec. 62, in which

$$\begin{aligned} f_1 &= 1,000 \text{ cycles} \\ f_{1\infty} &= 800 \text{ cycles} \\ R &= 1,000 \text{ ohms} \\ m &= 0.6 \\ C_1 &= 0.1325 \times 10^{-6} \text{ farad} \\ L_2 &= 0.1325 \text{ henry} \\ C_2 &= 0.298 \times 10^{-6} \text{ farad} \end{aligned}$$

From the considerations involving impedance matching at the end terminals, a mid-shunt termination facing each resistance termination is seen to be desirable for a series-derived section. Hence the structure of Fig. 80f is indicated. One full-series element (C_1) will be required, with a double-impedance shunt arm ($2L_2 + C_2/2$) at each end. The completed filter will then be as shown in Fig. 82.

2. Multisection Composite Filter. Required: Low-pass filter to be connected between resistance terminations of $R = 600$ ohms, with sharp cutoff at 1,000 cycles and high attenuation beyond.

There is no unique solution or "best" filter design for this problem. A large number of filters might be designed to meet these requirements, each of which would serve as well as any of the others. The relative merits of different designs will depend upon their economy of coils and condensers in accomplishing the required results. One suitable design is shown here:

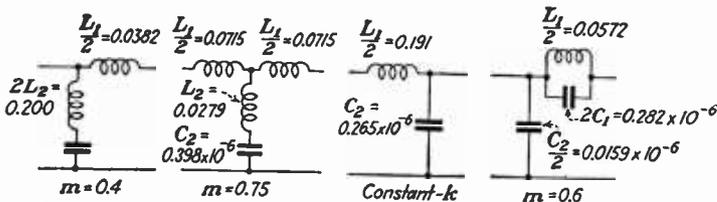


FIG. 83. Low-pass filter for use between 600 ohms with sharp cutoff at 1,000 cycles.

Let the input-end section be a half-section mid-series-derived m type, with its mid-shunt termination facing the input to match impedances at that point. Let $m = 0.4$ for this half section to give a sharp cutoff.

This will be followed by a symmetrical full section of the series-derived m type, mid-series terminated, with $m = 0.75$. Then a half section of the constant- K type with mid-series termination facing the full section and mid-shunt termination facing the end-terminating

half section, which will be shunt-derived m type, with $m = 0.6$. The latter will have a mid-shunt termination facing the constant- K half section and a mid-series termination facing the output termination.

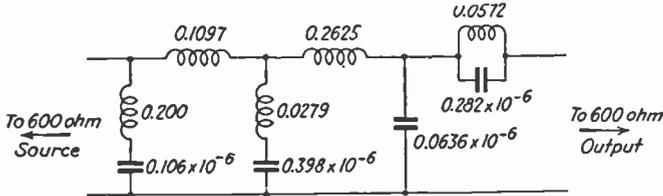


Fig. 84. Final filter as designed by Fig. 83.

References

Transients:

- Carter: "Simple Calculation of Electric Transients," Macmillan, 1945.
 Frank, E.: "Pulsed Linear Networks," McGraw-Hill, 1945.
 Kurtz, E. B., and G. F. Corcoran: "Introduction to Electric Transients," Wiley, 1935.
 Pierce, G. W.: "Electric Oscillations and Electric Waves," McGraw-Hill, 1920.
 Steinmetz: "Transient Electric Phenomena and Oscillations," McGraw-Hill, 1924.

A-c Circuits and Coupled Circuits:

- Aiken: Two-mesh Coupled Tuned Filters, *Proc. IRE*, February, 1937.
 Everitt, W. L.: "Communication Engineering," McGraw-Hill, 1937.
 Glasgow, R. S.: "Principles of Radio Engineering," McGraw-Hill, 1936.
 Morecroft, J. H.: "Principles of Radio Engineering," Wiley, 1921.
 Pierce, G. W.: "Electric Oscillations and Electric Waves," McGraw-Hill, 1920.
 "Radio Instruments and Measurements," *Nat. Bur. Standards Circ. 74*.
 Terman, F. E.: "Radio Engineering," McGraw-Hill, 1947.

Transmission Lines and Networks:

- Bode, H. W.: "Network Analysis and Feedback Amplifier Design," Van Nostrand, 1945.
 Bronwell, A. B., and R. E. Beam: "Theory and Application of Microwaves," McGraw-Hill, 1947.
 Brown: Characteristics of Unbalanced Overhead Transmission Lines, *RCA Broadcast News*, May, 1941.
 Duncan: Open Wire Transmission Lines, *Communications*, June, 1938.
 Everitt, W. L.: "Communication Engineering," McGraw-Hill, 1937.
 Guillemin, E. A.: "Communication Networks," Wiley, 1931, 1935.
 Hansell, C. W., and P. S. Carter: Frequency Control by Low Power Factor Line Circuits, *Proc. IRE*, April, 1936.
 Johnson: Ultra Low-loss Transmission Line for Television, *RCA Rev.*, June, 1946.
 Johnson, R. S.: "Transmission Circuits for Telephonic Communication," Van Nostrand, 1925.
 Kennelly, A. E.: "Electric Lines and Nets," McGraw-Hill, 1928.
 Mason, W. P., and R. A. Sykes: Use of Coaxial and Balanced Transmission Lines in Filters and Wide Band Transformers, *Bell System Tech. J.*, July, 1937.
 Meagher and Markley: "Practical Analysis of Ultra-high Frequency (Transmission Lines, etc.)," 2d ed., RCA Booklet, 1943.
 Paine: Graphical Solution of Transmission Lines, *Proc. IRE*, November, 1944.
 Palmer: Outline Notes on Telephone Transmission Theory, *Elec. Review (London)*, 1932.
 Pierce, G. W.: "Electric Oscillations and Electric Waves," McGraw-Hill, 1920.
 Selgin: Effect of Coil Q on Filter Performance, *Communications*, September, 1945.
 Slater, J. C.: "Microwave Transmission," McGraw-Hill, 1942.
 Sterba, E. J., and C. B. Feldman: Transmission Lines for Short Wave Radio Systems, *Bell System Tech. J.*, July, 1932.

Wave Filters:

- Campbell, G. A.: Physical Theory of Electric Wave Filters, *Bell System Tech. J.*, November, 1922.
 Carson, J. R., and O. J. Zobel: Transient Oscillations in Electric Wave Filters, *Bell System Tech. J.*, July, 1923.
 Johnson, K. S., and J. E. Shea: Mutual Inductance in Wave Filters, with an Introduction on Filter Design, *Bell System Tech. J.*, January, 1925.
 Johnson, K. S.: "Transmission Circuits for Telephonic Communication," Van Nostrand, 1925.
 Pierce, G. W.: "Electric Oscillations and Electric Waves," McGraw-Hill, 1920.
 Shea, T. E.: "Transmission Networks and Wave Filters," Van Nostrand, 1925.
 Zobel, O. J.: Theory and Design of Uniform and Composite Wave Filters, *Bell System Tech. J.*, January, 1923.
 Zobel, O. J.: Transmission Characteristics of Electric Wave Filters, *Bell System Tech. J.*, October, 1924.

See also cumulative indexes of *Bell System Tech. J.*, *Proc. IRE*, and *RCA Rev.* for additional references.

CHAPTER 6

ELECTRICAL MEASUREMENTS

BY R. F. FIELD¹ AND JOHN H. MILLER²

STANDARDS

1. General. True basic measurements of electrical quantities are rarely made except in standardizing laboratories, owing to the inherent difficulties in the procedure. Ordinary measurements are made by comparison devices of one form or another. Direct-reading instruments, having an electrical torque-producing means functioning against a spring, are calibrated against accurate standards which are in turn calibrated against basic measuring devices. Such torque-producing instruments are used for measuring current, voltage, power, and resistance. Instruments for measuring phase relations, frequencies, and other factors may have two torque-producing systems, each torque varying with the position of the moving element and bearing different functional relations to the quantity measured. The result is for the moving system carrying the pointer to take up a position where the torques balance, this being different for each different value of the quantity in question, and the scale may be marked accordingly.

Electrical units such as the volt, the ampere, and the ohm, are based upon and are intended to be exact multiples of the units of the centimeter-gram-second electromagnetic system. The international units which were made standard in 1911 were so derived. However, refinement of measurement techniques in the years following, along with intercomparison of the units between the national laboratories of various countries indicated that the international volt and ohm were slightly larger than the corresponding absolute values. Accordingly, as of Jan. 1, 1948, the absolute values were made standard.

The changes from the U.S. international values to the absolute values are minor, are tabulated below, and have been made to bring the electrical values into consistency with the absolute values of the fundamental mechanical units.

Values of U.S. International Units Legal to Dec. 31, 1947 in Terms of the Absolute Units, Legal after Jan. 1, 1948

1 international ohm	= 1.000495 absolute ohms
1 international volt	= 1.000330 absolute volts
1 international ampere	= 0.999835 absolute ampere
1 international coulomb	= 0.999835 absolute coulomb
1 international henry	= 1.000495 absolute henrys
1 international farad	= 0.999505 absolute farad
1 international watt	= 1.000165 absolute watts
1 international joule	= 1.000165 absolute joules

2. Current. Current is measured, absolutely, in terms of the force of attraction or repulsion between two coils connected in series and carrying that current, and the various dimensions of the coils. This current is then used to deposit silver in the *silver voltammeter* to determine the electrochemical equivalent of silver. One ampere of continuous unvarying current will deposit 0.001118 g of silver per second when following the standard procedure. The silver voltammeter is thus the standard of current.

¹ General Radio Company, Cambridge, Mass.

² Weston Electrical Instrument Company, Newark, N.J.

The use of this standard is tedious and time consuming, and it is generally used only for the exact calibration of a standard cell and a known resistance.

The standard of current is maintained, however, through the application of Ohm's law and the standards of resistance and voltage defined below.

3. Resistance. Resistance is measured absolutely by a number of methods in terms of a speed of revolution of a disk or coil and its various dimensions. The resistance is then compared with a mercury column of uniform cross section by a suitable bridge method. Such a column of mercury, having a mass of 14.4521 g, a uniform cross section (practically equivalent to 1 sq mm) of a length of 106.3 cm, and at the temperature of melting ice, has a resistance of 1 ohm.

The standard ohm is now maintained by the National Bureau of Standards as the average of a group of 1-ohm manganin resistors that have shown a change of less than 1 part per million from the mean of the group over many years. Practical secondary standards are coils of manganin wire immersed in oil and sealed in metal containers. Such sealed standards built by Leeds & Northrup Company to the specifications of the National Bureau of Standards are adjusted to an accuracy of 0.01 per cent and may be relied upon to hold their calibration to 1 part in 100,000 for considerable periods of time. The sealing of the containers is important to prevent the absorption, by the oil, of moisture from the atmosphere, for such moisture will deposit upon the shellac or other insulating material on the wire which, in turn, will cause mechanical strains to distort the values beyond normal expectancy.

4. Voltage. Voltage measurements cannot be made absolutely with an accuracy sufficient to make the measurement desirable, on account of the smallness of the electrostatic forces involved. The standard of voltage is maintained by the National Bureau of Standards as the average of a number of Weston *saturated* cadmium cells, sometimes known as the "Weston normal cell." The average value of this bank of cells is defined as 1.018300 international volts, or 1.018636 absolute volts at 20°C. These cells, as built by Weston and Epley, are correct to 0.001 per cent of the values given. The cell has a small but appreciable temperature coefficient, and in use its temperature must be maintained constant at the specified value. For this reason the saturated cell is not well adapted for commercial laboratory use.

The *unsaturated* cadmium cell, the type of standard cell customarily used for laboratory work, has a negligible temperature coefficient but must be compared with the saturated type for its initial calibration. Its voltage is constant to better than 1 part in 10,000, but the cells must be recertified yearly as the potential tends to drop very approximately 50 μ v per year. Standard cells must always be used in a null fashion, with no more than 50 μ a drawn from the cell at any time.

5. Reactance. The self and mutual inductance of single-layer air-core coils and the capacitance of two-plate air capacitors having guard rings may be calculated from their dimensions, with an accuracy of better than 2 parts in 100,000. The standard of capacitance is maintained by the National Bureau of Standards as a bank of quartz-insulated air capacitors.

6. Frequency. The absolute standard of frequency is the mean solar day as measured by astronomical observations. The mechanical vibrations of piezoelectric quartz crystals, or of tuning forks made from carefully stabilized metals, provide standards of frequency when permanently connected into suitable vacuum-tube circuits and allowed to oscillate continuously at constant temperature. Their frequency is constant to 1 part in 10 million per month and 5 parts in 100 million per day. Expressed in terms of time the latter figure is equivalent to 4 msec per day. Even with the best control of temperature and voltage there are short-time fluctuations of the order of 1 part in 1,000 million. The frequency of the crystal with which such accuracy may be attained is restricted to the neighborhood of 100 kc. Tuning-fork standards usually operate at 1,000 cycles. By means of suitable frequency multipliers and dividers all other frequencies from 1 cycle to 100 Mc may be obtained with the same accuracy.

Quartz crystals whose frequencies remain constant to 5 parts in 1,000,000 may be

made for the frequency range 20 kc to 10 Mc. Metals, such as nickel and certain iron alloys, having the property of magnetostriction, may be used as oscillators in suitable vacuum-tube circuits. Their frequency range extends from 5 to 100 kc. Their stability is about 2 parts in 100,000. For the lower frequencies tuning forks and metal bars are used. Their frequency range is 25 to 1,000 cycles.

CURRENT-MEASURING INSTRUMENTS

7. **Moving-coil permanent-magnet instruments** of the pointer type, or reflecting galvanometers, consist of a coil, usually wound on a metal frame for damping purposes, which can rotate in an intense uniform magnetic field produced by a permanent magnet.

The current I flowing through the turns N of the coil reacts with the magnetic field H in the air gap to produce a force F acting on each conductor proportional to the product IHl of the current, magnetic field, and length of conductor in the field. If the coil is pivoted at its center, a torque will be exerted, tending to rotate the coil about an axis parallel to the sides of the coil and perpendicular to the magnetic field. Some kind of restoring torque is provided which is proportional to the angle θ through which the coil rotates. Expressing the sensitivity S of the instrument as the angular deflection per unit current, it is given by



$$S = \frac{\theta}{I} = \frac{HNlb}{\tau} \quad (1)$$

FIG. 1. Moving-coil galvanometer.

where b is the diameter of the coil and τ is the restoring torque per unit angular displacement.

For maximum sensitivity a high flux density is desirable, although the effective upper limit appears to be about 4,000 gausses; higher values may act on otherwise imperceptible magnetic inclusions in the moving system and cause additional random torques which will give nonlinear deflectional characteristics. Further, since damping is a function of the square of the flux, high flux values may give overdamped deflectional characteristics even without a damping frame to the point where circuit damping alone may require many seconds for a final reading. Sensitivity increase by flux increase alone is, therefore, of limited usefulness.

8. **Torque to Weight Ratio.** The torque should be as low as possible for high sensitivity. Suspended-type instruments show no friction, but low torque lengthens the period. Again there is a practical lower limit. In pointer-type instruments where the moving element rotates on pivots between sapphire or the more recent high-silica glass V jewels, there is a minimum torque which may be used for a given moving element weight in order that frictional effects will be unobservable. For instruments mounted on a switchboard and having a horizontal axis, the ratio of the full-scale torque in milligram-centimeters with the weight in grams should not be less than 40 for small instruments, 60 for larger instruments of 1 per cent accuracy, and still greater if greater accuracy is required. For portable instruments having a vertical axis, it has been found that heavy elements, over 1 g, show greater friction than given by the above relation, and lighter elements show less friction. Hence for such vertical axis instruments for portable service the torque/weight^{3/4} ratio is used and this ratio should be over 40 for small instruments and over 60 to 100 for large instruments for unobservable friction. Ratios much lower than this may be satisfactory for highly sensitive laboratory instruments used with care and not subject to vibration or handling.

9. **Permanent Magnets.** The magnetic field obtained from the permanent magnet must be constant so that the electrical characteristics of the instrument may remain unchanged. The constancy of a magnetic system is determined by the ratio K , which is equal to the product of the effective length of the magnet times the effective cross section of one of the air gaps, divided by the product of the cross section of the magnet and the total air-gap length. This constant should be over 100 for chrome and

tungsten magnet steels and over 30 for high cobalt steels. For the various nickel-aluminum or MK steels the constant will vary, but 10 may be taken as a median value. Tungsten and chrome steels have been generally used in the past and still represent the most inexpensive type of magnet for moderate flux densities. High cobalt steels for higher gap flux values have been used extensively. However, for the higher values, and in new designs where the radically different characteristics can be effectively designed for, the various types of alnico are found most useful. Since the coercive values are very high, short magnetic systems are possible and, as a result, unorthodox designs are appearing. They usually tend to be self-shielded to a degree from external fields and appear to be very stable in use.

Flux densities in the air gap will run from 1,000 gausses up, with about 4,000 gausses representing the highest usable values. The structure of a pole piece and a core is used to decrease the length of the air gap and to make the magnetic flux uniform and radial. Where distorted d-c scales are required to balance other factors such as decibel relations, the pole tips may be cut away to produce a markedly distorted field resulting in a more uniform scale for the quantity measured.

10. Deflection Indicators. The deflection of any sensitive galvanometer is indicated by the angular rotation of a beam of light, the so-called *optical lever*, which is reflected from a mirror, either plane or convex, mounted above the moving coil. The older form of telescope and scale is now being replaced by a spot of light containing cross hairs which moves along a scale. The use of a spot of light is much less fatiguing than observation through a telescope, and a wider range of view is obtained. The usual scale length is 50 cm with zero in the center. The standard distance from mirror to scale is 1 meter. The maximum angular deflection is about 14 deg. Practically all pivot instruments use pointers. Full-scale deflection corresponds to approximately 90 deg. This is increased to 120 deg in some central-station meters by careful shaping of the pole pieces. It may be increased to 270 deg by a radical change in design.

11. Period and Damping. The moving element of every deflection instrument provided with a restoring torque proportional to the angular deflection is in effect a torsional pendulum. As such it has a moment of inertia P , a period T , and a damping factor. If the damping factor is low, the instrument will oscillate several times about its position of rest, each oscillation being less than the preceding one in accordance with the decrement of the system. For most rapid indication it is desirable that the instrument be not quite aperiodic or deadbeat but rather that it overswing from 3 to 5 per cent. (For a complete discussion of this see Drysdale and Jolley, "Electrical Measuring Instruments," Vol. 1, Chap. 3, Conditions for Rapid Indication, Ernest Benn, Ltd., 1924.)

Normal ammeters and voltmeters may be expected to have a period of the order of 1 to 2 sec. The smaller instruments, if equipped with magnets for very high gap densities and extremely light moving elements, may have a period as short as 0.2 sec (Weston high-speed power-level indicators). Instruments of ultrahigh sensitivity, where very little energy is available, may have a period as high as 5 sec. Sensitive suspension galvanometers may have a period as long as 12 sec.

The period of an instrument is important because the time necessary for any deflection instrument to attain a new position when its deflecting force is altered cannot be less than its period. High-speed indication in indicating instruments is very desirable, particularly when the phenomena being observed are rapidly changing, as in the monitoring of voice-frequency circuits; instruments with a long period will integrate the energy while high-speed instruments will give indications of peaks.

The friction of the suspension and the surrounding air is not sufficient to prevent the moving coil oscillating back and forth about its equilibrium position when a deflecting force is applied. The amount of damping is measured by the rate at which the amplitude of the oscillations decreases. The ratio of any two successive swings is constant. The Napierian or hyperbolic logarithm of this ratio is called the *logarithmic decrement* of the instrument. The smallest amount of damping which will cause

the coil to come to rest with no oscillation whatever is called the *critical damping*, and the coil is said to be critically damped. Increasing the damping beyond this point increases the time necessary for the coil to come to rest and produces overdamping. The shortest time in which the coil can come within a given small distance of its position of rest occurs when the coil is slightly underdamped. It has a value of about 1.5 times the period of the coil. The extra damping necessary to critically damp a coil is usually obtained magnetically from the motion of the coil in the field of the permanent magnet, which sets up counter electromotive forces. The amount of damping produced by the current in the coil depends upon the total resistance of the coil and connected circuit. That resistance which produces critical damping is called the *critical damping resistance* (CDR). A galvanometer is usually so designed that its critical damping resistance is at least five times its coil resistance so that it may be shunted for critical damping without losing much sensitivity. All but the most sensitive pivot instruments are damped on open circuit by the current set up in the metal winding form, and resistance of the connected circuit has little effect on the damping.

12. The current sensitivity of any galvanometer varies directly as the number of turns on its moving coil and as the square of its period. For a given winding space

Table 1. Characteristics of D-c Galvanometers

Make	Type	$E, \mu\text{v}$	$I, \mu\text{a}$	T	Resistance, ohms		$W, \mu\mu\text{w}$
					Coil	CDR	
Suspended-coil type with mirror							
Leeds & Northrup.....	2285a	0.032	0.0027	7.5	12	37	0.00009
	2285b	0.046	0.0038	5	12	52	0.00017
	2285f	0.032	0.00004	20	800	71,000	0.0000013
	2290	0.008	0.00001	40	800	101,000	0.00000008
	2500b	0.25	0.0005	6	500	10,500	0.00012
	2500e	1.5	0.003	3	500	2,500	0.0045
	2500f	0.05	0.0001	14	500	14,500	0.000005
	2239a	1.7	0.014	8	115	10,000	0.022
	2239b	1.0	0.001	14	1,000	10,000	0.001
	2239f	1.6	0.0002	18	8,000	54,000	0.00032
Suspended-iron type with mirror							
Leeds & Northrup.....	2270	0.008	0.0002	5	40	0.0000016
Suspended-coil type with self-contained scale							
Leeds & Northrup.....	2400c	10	0.01	3	1,000	16,000	0.1
	2420c	25	0.025	3	1,000	16,000	0.62
	2310d	125	0.125	3.5	1,000	11,000	15.6
	2430a	0.5	0.03	3.2	17	50	0.015
	2430c	0.13	0.005	2.5	25	400	0.0006
	2430d	0.27	0.0005	3.3	550	25,000	0.00014
Double-pivot type with pointer and scale							
Weston.....	440	37.5	0.25	2.7	150	1,150	9.4
	440	200	0.05	2.7	4,000	60,000	10.0

Values of voltage E , current I , and power W are for a scale deflection of 1 mm at a scale distance of 1 m for the galvanometers having mirrors; for those having self-contained scales the values given are for a deflection of the smallest division, usually 1 mm. The voltage drop in the external critical damping resistance is not included in the voltage given.

on the coil, its resistance varies as the square of the number of turns, assuming that the portion of the winding space occupied by insulation remains constant. The deflection is proportional to the current and to the square root of the resistance, *i.e.*, to the square root of the power dissipated in the coil.

In the selection of galvanometers it should be noted that in general those of high sensitivity will also be slow in action, and in general the natural period and critical damping resistance for a galvanometer as listed by the several makers should be considered as carefully as the sensitivity. Further, galvanometers of highest sensitivity require great care in leveling; they are responsive to minor vibrations and in many installations may require special supports.

13. Galvanometer Supports. Where vibration in a building is a factor, the Julius suspension may be used, a somewhat complex system of weights supported by springs with oil-damping vessels. A simpler method although not so perfect is to rest a 200-lb block (of concrete) on an air cushion; this will absorb all vibration usually encountered in factories, at least for galvanometers of moderate sensitivity. Galvanometers with a single suspension have the greatest sensitivity, those with a taut suspension less, and those with double pivots least. For the most sensitive type of galvanometer, increasing the period from 5 to 40 sec allows the power to be decreased from 11 to 0.005 μW . The minimum current sensitivity is 10^{-11} amp per mm. The smallest current sensitivity for a taut suspension is 10^{-8} amp per mm, and for a double-pivot pointer instrument, 5×10^{-8} amp per scale division.

14. Differential Galvanometer. Galvanometers of the suspended type are used mainly as null indicators for d-c bridges and potentiometers and as deflection instruments in comparison methods. In the latter case a *differential galvanometer* is sometimes used. This is a galvanometer having two separate insulated windings on the suspended coil. They have equal numbers of turns and are so connected that, when equal currents flow through the two coils, no deflection is produced.

15. Universal Shunt. The sensitivity of a galvanometer is most easily reduced by shunting, and, since it is desirable to keep the galvanometer critically damped, the Ayrton-Mather universal shunt shown in Fig. 2 is most convenient. This arrangement is also used in multiple-range ammeters and milliammeters and is frequently known as a "series shunt." The total resistance of the shunt is made approximately equal to the critical damping resistance of the galvanometer or indicating instrument with which it is used.

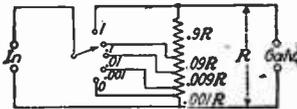


FIG. 2. Ayrton-Mather universal shunt.

16. D-c Ammeters and Voltmeters. Pointer-type instruments of the pivot type are used as ammeters and voltmeters of all ranges and as the indicating portions of thermocouple, rectifier, and various vacuum-tube instruments. The minimum range of the ammeters extends from 5 μa to an upper limit determined only by the size of shunt desired, commercial shunts having been made to 50,000 amp. Above 15 to 30 ma the movements are shunted, in which case the copper or aluminum winding of the moving coil must have sufficient manganin swamping resistance in series with it to give a good temperature coefficient when shunted by the manganin resistance. Voltmeters may be made with a full-scale range from 1 mv to as high as series resistance can be arranged to care for the requirements. Instruments are made with self-contained series resistance up to a few hundred volts; higher ranges usually require an external resistor with the instrument placed in the grounded or low-potential side of the circuit for the sake of safety and to reduce electrostatic effects on the moving system.

Voltmeter sensitivity is very commonly 1 ma for full scale, although laboratory standards of certain types may take as much as 100 ma in the interest of a balanced design for high precision. Conversely, for analyzing electronic circuits where the networks are of high resistance, a full-scale value of 50 μa is quite common, giving 20,000 ohms per volt.

While series resistors for low-range voltmeters are of conventional spool type, for ranges of over 1,000 volts tubular-type units are widely used, having resistance spools of special design, electrostatically shielded in sections contained in insulation tubes and filled with inert wax. Such units are completely moistureproof and mechanically well protected and are almost universally used for the measurement of plate potentials.

In general, pointer-type indicating instruments can be made to give full-scale deflection on as little as $0.1 \mu\text{v}$, although for a rugged instrument from 1 to $5 \mu\text{v}$ is required. Moving-element resistances may be made from about 1 to 10,000 ohms. Low-resistance elements are limited by the spring or suspension resistance which

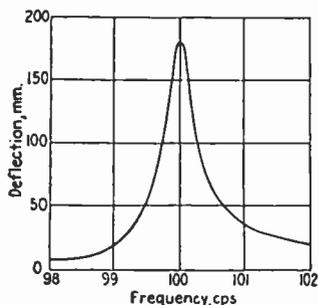


FIG. 3. Resonance curve of vibration galvanometer.

becomes a very appreciable part of the total, reducing the energy available for torque; high-resistance elements are limited by the available wire, and many are wound of enameled copper wire 0.001 in. in diameter.

As in the output circuits of vacuum-tube amplifiers, the resistance of the instrument or galvanometer should be matched to the circuit in which it is placed for maximum energy transfer, and this is particularly important where the energy is limited. On the other hand, this will frequently result in over damping galvanometers of ultrahigh sensitivity, and a compromise must usually be made between speed of response and sensitivity requirements. It should be noted, however, that this matching is not of vital importance since the loss by a very approximate match in error by as much as 20 per cent is very small.

17. Moving-coil Vibration Galvanometers. When an alternating voltage is applied to the coil of a permanent magnet galvanometer, the coil will follow the alternations of the current if the frequency is of the same order as that defined by its period. Maximum amplitude of vibration will occur at the natural frequency of the coil. The relation between amplitude and frequency is similar to the resonance curve of an electrical circuit. The ratio of the maximum amplitude at its natural frequency to the amplitude for an equal d-c voltage is between 25 and 150. The period of the ordinary d-c galvanometer is never less than 1 sec, while the frequencies at which measurements are made are rarely less than 30 cycles. The upper limit for a taut single suspension is around 300 cycles. This limit may be raised to 1,000 by the use of a taut bifilar suspension. Electrical characteristics of commercial vibration galvanometers are given in Table 2. At 60 cycles their sensitivity is equal to that of a good d-c galvanometer. A resonance curve when tuned to a frequency of 100 cycles is shown in Fig. 3.

The natural frequency may be raised still further by eliminating the coil entirely and using the single-turn loop formed by the bifilar suspension. The mirror is then placed at the center of the taut wires. The general method of construction is shown in Fig. 4. By this means a natural frequency of 12 kc may be obtained. The sensitivity decreases inversely as the first power of the frequency. On this account it is as sensitive at 10 kc as the bifilar-coil galvanometer was at 1 kc. In comparison with other null detectors at these frequencies, its sensitivity is so low that it is not much used in this form.

18. The Einthoven string galvanometer uses the simplest possible moving system for a galvanometer. A single conducting string moves in the narrow air gap of the magnetic system, which may be a permanent magnet or an electromagnet depending on the sensitivity desired. Its motion is observed through a microscope or by its shadow thrown on a screen from a point light source. The Einthoven string galvanometer built by the Cambridge Instrument Company uses a gilt glass fiber of

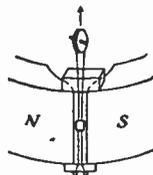


FIG. 4. Bifilar suspension.

0.002, 0.003, or 0.005 mm diameter with a resistance of 6,000, 4,000, or 1,400 ohms, has a period of 0.0036 to 0.01 sec, and at 1 m will deliver deflections of 2 to 200 mm per μa . The string galvanometer may also be used as an oscillograph. The shadow of the string is observed on a translucent screen as reflected from a revolving mirror. The motion of the string may also be photographed on film or bromide paper. The usual paper speed is 10 in. per sec, but this may be increased to a maximum of 100 in. per sec. At this latter speed, phenomena lasting 1 msec appear 0.1 in. long.

Table 2. Characteristics of A-c Galvanometers

Make	Type	f , cycles	E , μv	I , μa	R , ohms	W , $\mu\mu\text{w}$
Leeds & Northrup.....	2350a	60	17.5	0.025	700	0.44
Suspended-coil type with electromagnet						
Leeds & Northrup.....	2570	60	0.06	0.005	12	0.0003
Electromagnet type						
Leeds & Northrup.....	2440	60	16	0.05	325	800,000
Vibrating-diaphragm type (telephone)						
Western Elec.....	800	400	0.02	6,000	2.4

Values of voltage E , current I , and power W are for a scale deflection of 1 mm at a scale distance of 1 m for all galvanometers except the telephone, for which the threshold of audibility is used. The moving system is tuned to the frequencies given for all instruments except the suspended-coil galvanometer with electromagnet.

19. **Moving-coil A-c Instruments.** If a steady deflection is desired with a.c., the magnetic field must change in direction with the current in the coil and must have the same phase. This requires that the field be an electromagnetic one. In the case of galvanometers and particularly null indicators, a field of laminated iron may be used, excited at the same frequency as the moving coil. When used as a null indicator in a bridge network, the field is connected across the same supply as the bridge, while the moving coil is connected to the detector terminals. Since the current through the field and the flux produced will be nearly 90 deg out of phase with the voltage applied to the bridge, the galvanometer will be most sensitive to the reactance balance and will be little affected by the resistance balance. These conditions may be equalized or reversed by the introduction of resistance in series with the field, or reactance in series with the bridge, to make the field current and bridge current differ in phase by 45 deg or be in phase. The phase selectivity of the a-c galvanometer may be of advantage in certain special cases, but in general it is a considerable disadvantage. The electrostatic field of the main field winding exerts a considerable force on the moving coil so that it must be carefully shielded. The sensitivity of the a-c galvanometer is very high and it compares favorably with the best d-c galvanometers.

20. **Electrodynamometer.** When the iron core is omitted from the field winding, the moving coil and field coil may be connected in series. The deflection is then proportional to the square of the current flowing in the windings, and the instrument is called an *electrodynamometer*. Instruments of this type read the same on both a.c. and d.c. and are suitable as transfer instruments, provided certain precautions are taken. Protection from external magnetic fields is most important. This is usually accomplished in pivot-type instruments by shielding with soft iron. It may also be effected by making the instrument *astatic*. When a.c. is used, an error is introduced if the distribution of current in the coils is affected by eddy currents in the conductors

themselves—the so-called *skin effect*—or by capacitance between windings. The former effect is minimized by the use of conductors with insulated strands—so-called *litzendraht*—the latter by careful spacing and by electrostatic shielding.

Electrodynamometers may be used as galvanometers, ammeters, voltmeters, and wattmeters. Their sensitivity as galvanometers is so low compared with vibration galvanometers and other meters that they are now rarely used. As ammeters, voltmeters, and wattmeters, they are the standard instruments for use at commercial frequencies. In general the sensitivity of a-c instruments is of the order of 1/1,000 of that of d-c instruments, this being due to the difference in field intensity of the electromagnetic field as compared with that which can be obtained from a permanent magnet. Electrodynamometer instruments of the highest precision will take from 1 to 3 watts full scale, the total energy varying with the square of the deflection. Suspension-type electrodynamometers may have sensitivities 100 times as great.

Table 3. Characteristics of A-c Ammeters

Make	Type	<i>E</i> , volts	<i>I</i> , amp	<i>R</i> , ohms	<i>W</i> , w
Electrodynamometer type					
Weston.....	326	2.6	1.0	2.6	2.6
	341	1.0	0.5	2.0	0.5
	370	21	0.015	1,400	0.31
Moving-iron type					
Weston.....	155	31	0.02	1,540	0.62
	433	14	0.03	400	0.41
	476	30	0.015	2,000	0.45
	517	30	0.015	2,000	0.45
	528	30	0.015	2,000	0.45
Thermocouple type					
Cambridge.....	0.010	30	
	0.10	4	
	1.0	0.2	
	622	0.25	0.01	25	0.0025
Weston.....	425	0.13	0.10	1.35	0.0135
	425	0.62	0.12	5.2	0.075
	425	0.59	0.50	1.18	0.295
Rectifier type					
Weston.....	301	1	0.001	1,000	0.001
Westinghouse.....	PY-4	13.4	0.010	1.270	0.13
	NA	13.6	0.010	1.300	0.13

Values of voltage *E*, current *I*, and power *W* are for full-scale deflection.

Electrodynamometer ammeters have their fields and moving coils in series up to several hundred milliamperes above which the moving element is shunted across a resistor in series with the fixed coils. Above 50 amp, or so, current transformers are used, and these are now available with special alloy cores which will give accuracies of the order of $\frac{1}{10}$ of 1 per cent. Electrodynamometer instruments are ordinarily made to function up to 125 cycles without correction but may be used on frequencies up to several thousand cycles if especially designed or if corrections are made. Note

that low-range voltmeters have very low resistance in order to get the required energy; dynamometer voltmeters with full-scale values of 2 volts may draw as much as 0.5 amp. High voltages above 1,000 volts are measured with potential transformers.

Electrodynamometer instruments are also used as wattmeters where the field is excited in series with the load and the moving coil is across the load in series with suitable resistance, the readings being proportional to $EI \cos \theta$. For polyphase circuits a multiplicity of similar elements may be arranged on a single shaft, the most usual variety being the two-element instrument on three-phase circuits. Such an instrument gives true power without relation to phase angle.

21. Moving-iron Instruments. Galvanometers may be constructed with a stationary coil and a moving-iron vane or magnet. The moving system consists of small permanent magnets placed at the center of the coil at right angles to the axis of suspension. To avoid the effect of outside magnetic fields, the system is duplicated with the magnets pointing in the opposite direction to make it astatic, and the whole galvanometer is surrounded by multiple soft-iron shields. Its sensitivity (see Table 1) is nearly equalled by the best moving-coil galvanometers so that it is very little used.

Soft iron may also be used in the moving element, either alone or in conjunction with a fixed piece of soft iron, both of which are magnetized by the fixed coil.

Soft-iron meters are much used as a-c ammeters and voltmeters in a wide variety of ranges and sizes. They may also be used on d.c. Electrical characteristics are given in Table 3. The range of the ammeters is from 20 ma to 500 amp. The upper limit is ten times that of dynamometer-type meters, because the current coil is fixed. Currents up to 5,000 amp are measured by the use of current transformers. Frequencies to 500 cycles may be used. The range of the voltmeters is from 1 to 750 volts. Their resistances are such as to give from 3 to 200 ohms per volt, the values increasing with the voltage. Higher voltages are measured by the use of either multipliers or potential transformers. Frequencies up to 500 cycles may be used, the normal limit being 125.

In general the sensitivity of pointer-type indicating instruments using the moving-iron principle is from 0.1 to 1 watt full scale. Instruments using short vanes, usually of the arcuate type, take about 1 watt full scale. Instruments with long radial vanes are more sensitive with a minimum of 0.1 watt full scale but in general are more sensitive to external fields and must be well shielded and kept away from strong external fields. Moving-iron instruments in general are less satisfactory on badly distorted wave forms as the hysteresis loop of the iron is represented in the measurement. They are, however, widely used on power circuits and are generally available in all sizes from the small 2-in. instruments up to the larger switchboard types.

HIGH-FREQUENCY CURRENT METERS

22. To measure currents of high frequency, the only satisfactory means is through the heat developed in a resistor, which heat may be measured by the expansion of a wire, by measuring the thermoelectric voltage developed by a thermocouple adjacent to the resistor wire, by bolometer methods, and by other heat-measuring systems.

23. The hot-wire expansion type of instrument is today practically obsolete. Its defects of varying in indication with ambient temperature, the lack of perfect resiliency in the heated expansion wire, and its low overload capacity together with the advent of the thermocouple instrument have practically made this type obsolete.

24. A thermocouple meter consists of a heater member, a thermocouple adjacent to it, and a d-c galvanometer or millivoltmeter. Figure 5 shows the basic diagram of the device. Such a simple assembly, however, does not compensate for variations in temperature of the terminals or for ambient temperature variations.

The Weston thermal ammeter as developed by W. N. Goodwin, Jr., is as shown in Fig. 6. The heater is a wire or tube of platinum alloy of very short length whereby most of the heat is conducted to the terminals, thus wiping out largely the effect of convection currents of air. The temperature of the heated member may be represented as a parabola in its gradient from center to each terminal lug, and it is this

temperature difference or gradient from the center of the heater to its end which is measured by the thermocouple. The couple proper consists of a pair of wires, usually of constantan and a platinum alloy, permanently welded to the center of the heater at the junction end, with the effective cold ends soldered to a pair of copper strips which are thermally connected to, but electrically insulated from, the terminal lugs. Their heat capacity is such that the difference in temperature between the center of the heated member and the center of the two copper compensating strips is always the same as from the center of the heated member to the terminal lug, regardless of

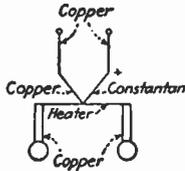


FIG. 5. Thermocouple meter.

ambient temperature changes or general rise in temperature of the surroundings due to heating of the lugs themselves or temperature rise due to the total heat generated. The thermoelectric voltage is, therefore, strictly proportional to the temperature difference between the center and ends of the heated member which in turn is proportional to the square of the current causing this temperature rise, and a d-c instrument connected to the couple may be calibrated in terms of this current. Couples may be designed to give suitable indication on instruments of commercial types from 200 ma up to whatever may be required. Solid round wires may be used for the heated member up to about 2 amp, but for higher currents and at the higher frequencies skin-effect phenomena cause the readings to be too high. For higher ranges, therefore, the heated member should preferably take the form of a thin-walled tube of such dimensions that the frequency error will not be excessive. For a frequency error of 1 per cent at 50 Mc, and for a platinum alloy tube having a resistivity of 170 ohms per cir mil ft, the wall thickness is about 0.001 in.; at 100 Mc the readings will be about 5 per cent high. Since a 1-mil wall represents a practical minimum thickness of the heated tubular element, thermoammeters are not useful for even moderate accuracy at frequencies much higher than 150 Mc; as a practical matter above 150 Mc standing waves of even minor magnitude preclude accurate measurement of current values as such, and the introduction of a line discontinuity by adding the thermoelement may also cause gross errors. Current measurements in the usual sense are, therefore, rarely made at above 150 Mc. It should be noted, however, that as long as the pointer of a thermoammeter does not deflect beyond full scale, no damage will occur to the thermoelement itself at any high frequency, and qualitative measurements may be made on frequencies much higher than 150 Mc.

While standard instruments have a square-law scale as the result of the I^2R production of heat, instruments are available in which the upper four-fifths of the scale is approximately linear through the use of special d-c indicating mechanisms having nonlinear air gaps whereby the d-c sensitivity is progressively lower as the pointer moves up the scale. By a proper combination of such specially shaped pole pieces a nearly linear scale may be produced. (See Figs. 15c and d.)

Instruments having the linear expanded scale are useful in small broadcast transmitters licensed for a lower power at night than during the day; sufficiently accurate readings of the high and low values of antenna current may be had on the same instrument to be satisfactory, and instruments of this type are listed as complying with FCC rule No. 143.

For low ranges so-called *bridge-type* couples are used, as shown in Fig. 7, whereby

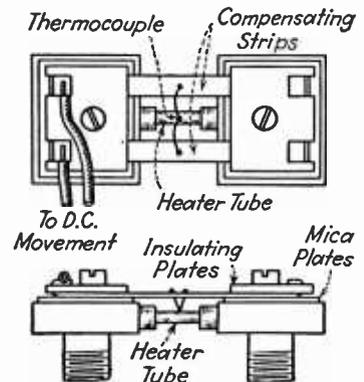


FIG. 6. Compensated high-frequency thermocouple and heating element.

a number of couples are arranged in series-parallel to give a higher thermal emf. The impedance of these couples is higher than for a single couple, and for the common current-squared galvanometer the effective resistance is 4.5 ohms. The indicating instrument for the standard single couples has a sensitivity of 12 mv and a resistance of about 5 ohms.

For still higher sensitivities the couple may be placed *in vacuo*. Such couples show no increase in sensitivity until the vacuum is better than 0.01 mm. of mercury; as the vacuum is increased, the sensitivity of the low ranges with very fine heaters increases markedly since there is no cooling by convection currents. Characteristics of a line of vacuum thermoelements are listed in Table 4, as made by the Weston Electrical Instrument Corp., and are typical of those broadly available. Metallic heaters are used on the higher ranges, with a carbon filament for the heater of the 1.5 and 2.0 ma ratings. These thermoelements are to be used with 10-ohm, 200- μ a instruments, and are used either external to the instrument or self-contained in the instrument case. In some instances they are associated directly with the coaxial line to simplify the r-f system, with leads to the instrument. Frequency errors will vary with range; up to 150 Mc such errors are usually under a few per cent. Care must be taken to avoid r-f currents in the couple

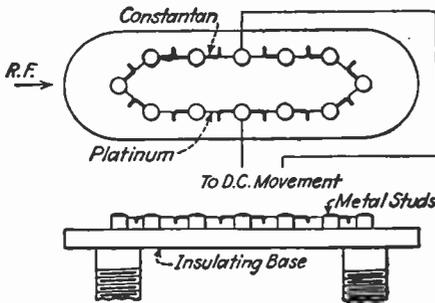


Fig. 7. Galvanometer or bridge-type couple.

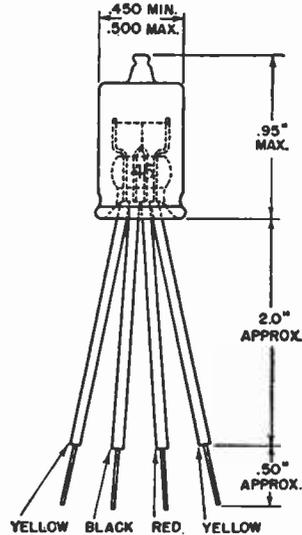


Fig. 8. Physical dimensions of thermoelements listed in Table 4.

which will heat the junction independently of the heater current; chokes in the instrument leads and capacitance to ground may be required to maintain this condition.

Thermocouple instruments in general are calibrated on commercial frequency a.c.; if used on d.c., the mean of reversed readings should be taken to make certain that any d-c drop in the heater picked up by the couple is canceled out.

Thermocouple instruments may be obtained with separate couples for use in indicating at a distance as where a couple is placed in the antenna of a transmitting station and the leads brought back to an instrument in the transmitting building. The couple should be placed in the h-f circuit at a point close to ground potential to reduce circulating r-f currents in the leads to the instrument. If this cannot be done, the thermocouple, of low range, around $\frac{1}{2}$ amp, is connected to a loop of wire that is inductively coupled to a loop in the main antenna circuit. The thermocouple circuit may then be grounded. The instrument is scaled to read the total antenna current, and the final calibration is made by adjusting the inductive coupling between the two loops until the remote reading instrument indicates the same value as an instrument placed directly in the antenna itself. Note that FCC rules require an instrument in the main antenna circuit which may be used for this purpose but which under normal operating conditions is kept short-circuited to prevent damage due to lightning. The

Table 4. Characteristics of Vacuum Thermoelements

Range, ma*	Resistance, ± 10 per cent		Max safe heater current, ma	Approx burn-out current, ma
	Heater, ohms†	Couple resistance, ohms		
1.5	1,365	6	3.2	20
2.0	750	6	5.0	30
5.0	82	6	10.0	20
7.5	36.2	6	16	35
10	23.4	6	25	50
15	13.0	6	40	85
20	8.4	6	50	120
25	7.0	6	62	150
30	5.8	3	75	180
37.5	4.6	3	85	210
50	3.3	3	115	260
75	1.36	3	170	330
100	1.03	3	220	380
150	0.66	3	320	540
200	0.46	3	420	700
250	0.39	3	510	900
300	0.33	3	610	1,050
400	0.25	3	800	1,300
500	0.20	3	1,000	1,900

* Heater current for 5.0 mv open circuit may be 20 per cent less than rated current.

† Resistance of heater at rated current.

switch is opened when the instrument is read for logging purposes, and the remote indicator, usually located on the transmitter panel, is used for normal operation.

The ratio of the power available to operate the indicating meter to that put into the heater is about 1 to 2,000 for the most efficient couples; hence a very sensitive d-c instrument is required for low r-f energies.

Thermocouple voltmeters are constructed by using one of the more sensitive couples with sufficient series resistance to give the desired voltage range. Their range is from 0.3 to 150 volts with resistances of 125 ohms per volt above 1 volt, and 500 ohms per volt above 10 volts, if desired. Their frequency range is determined by that of the series resistance. The small resistance spools which must be used in meters with self-contained resistors change their resistance rapidly with frequency so that their frequency limit is 3 kc. Frequencies of 1 Mc may be attained with an error of 1 per cent with special h-f resistors.

Since the emf produced by the thermocouple is proportional to the power input and hence to the square of the current, this meter will read correctly on both d.c. and a.c. and may, therefore, be used as a transfer instrument. It is necessary, however, to take the average of the readings for both directions when using d.c.

RECTIFIER METERS

25. An a.c. may be changed to a pulsating current having a steady component by the process of rectification. If the current-voltage characteristic is as shown in Fig. 9a the effect is called *half-wave rectification*. The negative half cycles are eliminated and the positive half cycles reproduced undistorted. The value of the steady component is half the average value of a half sine wave. The ratio of the d.c. to the effective value of an a-c current having a sine wave form which would flow if the rectifier were replaced by a pure resistance of the same value as that of the rectifier is $\sqrt{2/\pi}$, or 0.450. By a combination of rectifiers it is possible to obtain the characteristic shown

in Fig. 9b, which gives full-wave rectification. The d.c. is then 0.900 of the a.c. Actual rectifiers have a curved characteristic as shown by the dotted line in Fig. 9a. For negative voltages the resistance is not infinite. The ratio of the positive and negative half-cycle resistances is sometimes as low as 8. Because of the curvature of the characteristic, the ratio of d.c. to a.c. is a function both of the magnitude of the current and of wave form.

The crystal rectifiers used with early radio receivers may be used with a sensitive d-c meter for rectifying an a.c. Carborundum, galena, silicon, germanium, and many other crystals may be used. The crystal is cast in a low melting-point alloy and the top contact made with a fine copper wire which rests on the crystal or a fine tungsten wire which is welded to the crystal. Rectification occurs at the points of contact of metal and crystal.

26. Commercial rectifier instruments contain a full-wave rectifier consisting of four copper oxide rectifier disks connected in bridge relation as shown in Fig. 10. The rectification is by virtue of the oxide film formed on the copper disk. Current flows readily from the oxide to the copper and much less readily in the reverse direction.

For instrument use the rectifier consists of four small plates arranged in a stack with suitable terminals between adjacent disks for connection to the instrument and the external circuit. The disks may be as large as $\frac{3}{16}$ in. square or round, which size is rated at about 1 volt and 5 ma maximum. This rating is somewhat less than a maximum rating for power purposes since in an instrument some overload capacity is required and stability rather than maximum power is the main requirement. Somewhat smaller disks are used in low-range instruments and for those designed for special characteristics in order to maintain a relatively high current density at lower currents, thus reducing frequency errors. Contact with the oxide is made in a variety of ways through the use of lead washers, graphite, or various metals applied to the surface. The main requirement here is permanence of contact over an extended period.

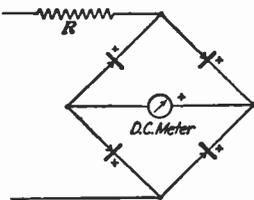


FIG. 10. Copper-oxide rectifier bridge.

The sensitivity¹ of the device depends upon the resistance and full-scale current of the d-c instrument. The d-c instrument measures the average value of a rectified wave, while a.c. is usually measured by methods which give the rms value of the wave. It is customary to calibrate rectifier instruments in terms of the rms value of a stated wave form, usually a sine wave. If a rectifier instrument is used on a wave form differing widely from the wave for which it is calibrated, an error proportional to the form factor will result. Calibration also corrects an error due to imperfect rectification, which varies with current, temperature, and frequency.

The performance of rectifier instruments can be best expressed by considering the d-c instrument and the rectifier as a unit according to Fig. 10. The current efficiency,

$$F = \frac{\text{average d-c current}}{\text{rms a-c current}}, \text{ is 80 to 89 per cent for a sinusoidal a-c current in the order}$$

of 0.001 amp. It is therefore impossible to use an a-c rectifier instrument for d.c. without first making a suitable change in circuit or calibration. Figure 11 shows the effect of current on current efficiency for a sinusoidal wave. This variation is corrected in calibrating.

The 60-cycle impedance of a 20-ma rectifier instrument is shown in Fig. 12. Other

¹ The following several paragraphs and Tables 5 and 6 have been contributed by F. S. Stickney of the Westinghouse Electric Corp.

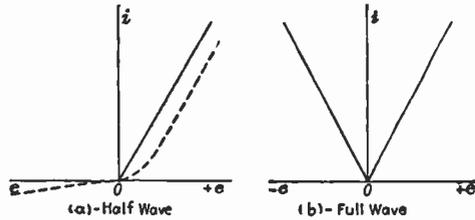


FIG. 9. Rectifier characteristics.

ranges using different rectifiers will have different values, but in general the slope of the characteristic as plotted in logarithmic coordinates will be entirely similar.

Temperature variations have considerable effect on both the impedance and accuracy of rectifier instruments. Figure 13 shows temperature-voltage variations for a specific group of milliammeters from which impedance can be determined.

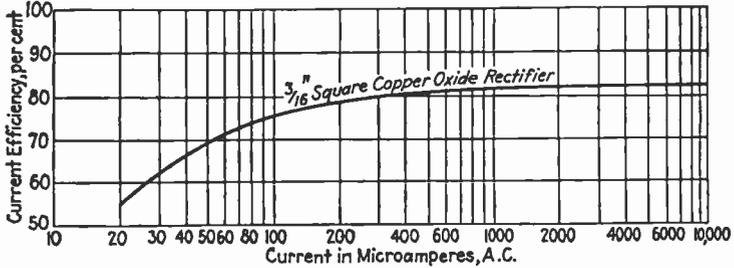


FIG. 11. Current-efficiency characteristic.

Figure 14 shows temperature-efficiency relations of this group at various current values. The point must be stressed, however, that the curvature of these characteristics varies with the several parameters of rectifier-disk size, current density, processing time, and the resistance of the instrument, and it is quite possible to modify these curves materially for special requirements. Standard instruments, by the same token, can hardly be represented by any particular group of curves. It might be stated that

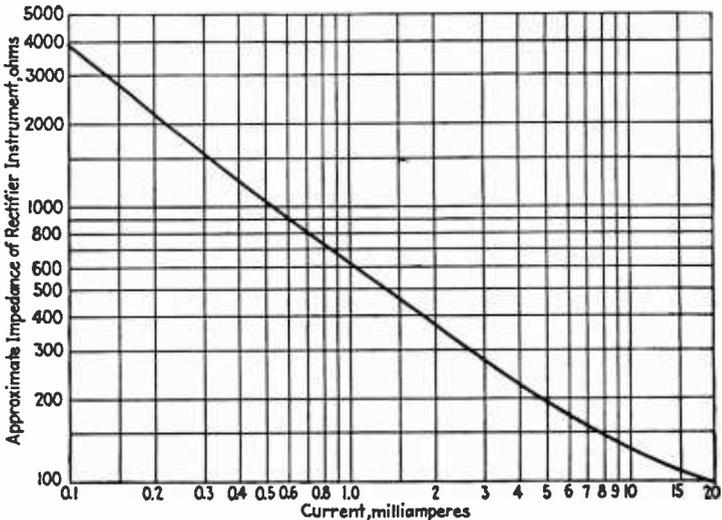


FIG. 12. Impedance characteristic of 20-ma rectifier instrument at 60 cycles.

rectifier instruments have been materially improved in recent years as to the flattening of the curves and that design possibilities have broadened to the point where materially improved instruments can be made for particular requirements.

Higher temperatures adversely affect the rectifying film, and rectifier instruments may become erratic at temperatures in excess of 45°C. High-temperature locations should be avoided in application; where the instrument becomes unduly warm, instru-

ments with external rectifiers are sometimes used with the rectifier placed in a relatively cool location.

Frequency errors are the result of capacitance between disks. Since the disk resistance is lower at higher currents and since capacitance is a function of rectifier size, the smallest rectifier is preferred for good frequency characteristics. This in turn means a high current density with which good accuracy is obtainable somewhat above audio frequencies. With low-current density, errors may be as large as 1 per cent per 1,000 cycles.

In general, low-range voltmeters are more subject to temperature and frequency errors than high-range voltmeters. Low-range voltmeters have scales which

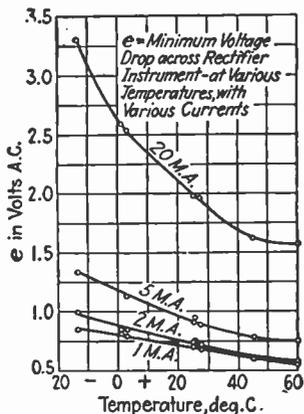


FIG. 13. Effect of ambient temperature on the voltage drop across a rectifier instrument at various currents.

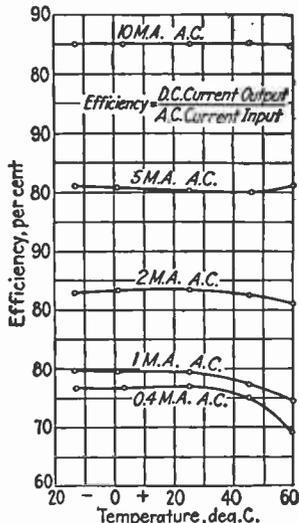


FIG. 14. Ambient temperature-efficiency relation of several typical rectifier-type milliammeters.

are compressed at the lower end due to variations of impedance with current. High-range voltmeters and milliammeters have nearly uniform scale distribution.

Tables 5 and 6 give approximate constants of commercial rectifier instruments.

Table 5. Milliammeters and Microammeters

Full Scale, Ma	Approximate 60-cycle Impedance at Full Scale*
15	100
10	130
5	190
2	370
1	600
0.5	1,140
0.2	1,950
0.1	4,200
0.05	6,300
0.02	10,000

* Individual copper oxide rectifiers vary considerably from the average in characteristics. Impedance values given may vary ± 15 per cent, and efficiency values vary ± 3 per cent for the product of one manufacturer. Much greater variations may be expected between the products of different manufacturers.

27. Power-level instruments used in the monitoring of voice-frequency circuits are usually voltmeters with scales calibrated to read power on the basis of a fixed-resistance load. The indications of power are usually in decibels above or below a specified zero power level. Prior to 1939 considerable confusion existed in this field of measurement owing to the fact that zero levels of 1, 6, and 12.5 mw were used into loads of 500 or

600 ohms. The instruments themselves, fundamentally voltmeters of the rectifier type, have been quite satisfactory. The usual impedance has been 5,000 ohms and higher to avoid too great a loss due to the addition of the power-level indicator and also to avoid adding harmonics to the line due to the non-linear shunt resistance of the instrument-rectifier network.

Table 6. Voltmeters

Full scale, volts	Full scale, approximate ohms per volt	Approximate fixed resistance, ohms	Approximate 60-cycle impedance of rectifier and d-c instrument at full scale, ohms
150	1,000	149,400	600
50	1,000	49,400	600
10	1,000	9,400	600
4	1,000	3,400	600
3	2,000	4,860	1,140
2	2,000	2,800	1,140
1.5	2,000	1,860	1,140
1	5,000	3,050	1,950
0.5	5,000	550	1,950

This situation has been largely cleared due to the work of Chinn, Gannett, and Morris¹ in the development of the VU meter. This is fundamentally a rectifier voltmeter having very definitely specified electrical and ballistic characteristics. These characteristics are specified in detail, along with a definition of the term VU, in "American Recommended Practice for Volume Measurements of Electrical Speech

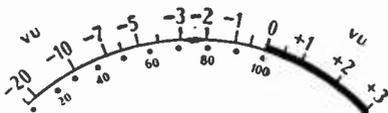


Fig. 15a. A scale for VU meter.

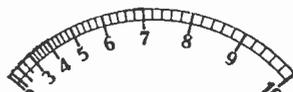


Fig. 15c. Standard scale using a conventional d-c movement.

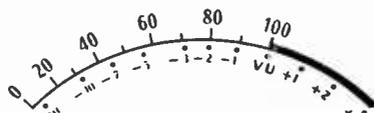


Fig. 15b. B scale for VU meter.

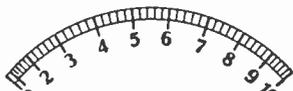


Fig. 15d. Linear expanded scale using the mechanism with specially shaped pole pieces.

and Program Waves," sponsored and published by IRE and known as American Standard C16.5-1942.

Two instrument scales have been standardized, as shown in Fig. 15. The upper scale, known as the *A scale*, emphasizes the VU markings and has an inconspicuous voltage scale. The lower, known as the *type B scale*, emphasizes the per cent voltage and has a relatively inconspicuous VU scale. This latter scale is largely used in broadcast monitoring since the voltage scale indicates in a rather direct fashion the per cent utilization of the facilities. The scales are printed on buff paper to reduce eyestrain;

¹ Chinn, Gannett, and Morris, *Proc. IRE*, January, 1940; A New Standard Volume Indicator and Reference Level, *Bell System Tech. J.*, January, 1940.

the narrow arc and the figures above it are in black with the heavy arc to the right, the markings above it as well as the markings below the arc in red.

The instrument mechanism, which is identical for both scales, has very definite ballistic characteristics which may be completely defined by the fact that, if a voice-frequency voltage of such amplitude as to give a steady reading of 100 on the voltage scale is suddenly applied, the pointer should reach 99 on this scale in 0.3 sec and should then overswing the 100 point by between 1 and 1.5 per cent.

Zero level is 1 mw in 600 ohms. Since a voice-frequency channel may contain many components of different frequencies and since they may affect different instruments in a different manner, the ballistic standards above listed are a very necessary

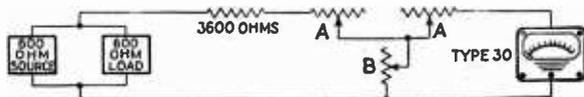


FIG. 16. Network for use with VU meter.

part of the new standard. The instrument is standardized on sine-wave voltage and is adjusted to read to the 100 mark on the voltage scale with 1.225 volts applied, this representing 4 db above 1 mw in 600 ohms and is applied to the standard instrument as furnished, plus a 3,600-ohm external series resistance.

With such an instrument, the readings obtained from it when voice-frequency currents are applied may then be stated as so many VU, taking into account that 4 VU must be added to the scale reading plus the number of VU lost in the attenuator placed in the network.

The required network is shown in Fig. 16. The fundamental total resistance of the instrument is 7,500 ohms. To this are added 300 ohms representing a 600-ohm

Table 7. Attenuators for VU Meter

Attenuator loss, db	Level, VU	Arm A, ohms	Arm B, ohms	Attenuator loss, db	Level, VU	Arm A, ohms	Arm B, ohms
0	+ 4	0	Open	24	+28	3,437	494.1
1	+ 5	224.3	33,801	25	+29	3,485	440.0
2	+ 6	447.1	16,788	26	+30	3,528	391.9
3	+ 7	666.9	11,070	27	+31	3,566	349.1
4	+ 8	882.5	8,177	28	+32	3,601	311.0
5	+ 9	1,093	6,415	29	+33	3,633	277.1
6	+10	1,296	5,221	30	+34	3,661	246.9
7	+11	1,492	4,352	31	+35	3,686	220.0
8	+12	1,679	3,690	32	+36	3,708	196.1
9	+13	1,857	3,166	33	+37	3,729	174.7
10	+14	2,026	2,741	34	+38	3,747	155.7
11	+15	2,185	2,388	35	+39	3,764	138.7
12	+16	2,334	2,091	36	+40	3,778	123.7
13	+17	2,473	1,838	37	+41	3,791	110.2
14	+18	2,603	1,621	38	+42	3,803	98.21
15	+19	2,722	1,432	39	+43	3,813	87.53
16	+20	2,833	1,268	40	+44	3,823	78.01
17	+21	2,935	1,124	41	+45	3,831	69.52
18	+22	3,028	997.8	42	+46	3,839	61.96
19	+23	3,113	886.3	43	+47	3,845	55.22
20	+24	3,191	787.8	44	+48	3,851	49.21
21	+25	3,262	700.8	45	+49	3,857	43.86
22	+26	3,326	623.5	46	+50	3,861	39.09
23	+27	3,384	555.0				

source, and load in parallel, making a total of 7,800 ohms. To simplify the use of an attenuator, this is split in the center to give 3,900 ohms each side, which will allow for a simple T-pad attenuator to be inserted at this point. The instrument proper, therefore, has an internal resistance of 3,900 ohms and must be used with the separate 3,600-ohm resistor. Since the normal instrument level is +4 VU, the attenuator dial is marked 4 VU at zero attenuation, and for other true attenuation values 4 VU are added. Table 7 shows values for such attenuators.

This instrument is available commercially and because of its deliberate action is found most readable. The standardization of the instrument by the majority of those concerned is of fundamental importance, particularly where levels along a transmission line are to be read, forwarded over an order wire to a common point, and compared. While instruments of several sizes are available, the one in most common use is approximately 4 in. square and available either with or without internal illumination.

The advent of this new level indicator has very largely superseded the use of db meters, with scales as shown in Fig. 17, although the previously available high-speed instruments still find some utility, particularly in the cutting of records where instantaneous indication and control of high level is necessary to prevent overcutting.

Table 7 is a useful tabulation of power levels, ratios, and voltages, all in terms of the zero level of 1 mw in 600 ohms, and, when interpreted on an instrument of the characteristics described, the values of db above and below this level will also represent VU.

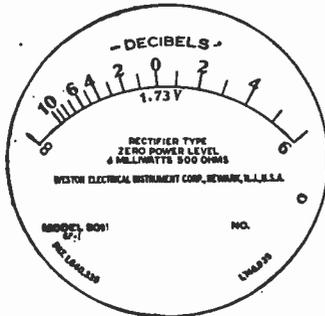


FIG. 17. Scale of db meter.

MEASUREMENTS OF PULSATING CURRENTS AND POTENTIALS

In making measurements of current and voltage which are neither true a.c. nor d.c., care must be taken to make the measurement with the correct type of instrument in order that a measurement be had of the actual value required.

28. Rectified current, which may or may not be filtered, should in general be measured with a moving-coil permanent-magnet type of d-c instrument. This gives the average value. It is the value of current or voltage of interest when charging a battery and in general is the value of interest in vacuum-tube technique. Iron-vane and electro-dynamometer instruments indicate the rms value which is used for determining the heating effect.

Direct-current instruments, particularly voltmeters, have a sufficiently large heat-overload capacity so that they may ordinarily be used on pulsating currents without danger.

To measure the a-c component of voltage, a capacitor may be placed in series with an a-c voltmeter of suitable range; the d-c component is blocked and the a-c value only is measured. The impedance of the capacitor at the frequency used (120 cycles for a full-wave rectifier system) should not be greater than 10 per cent of the instrument resistance; the impedances being in quadrature, the resulting error will be under 1 per cent. This is the simplest method of measuring hum in a rectified plate supply. Because of its high resistance, the rectifier voltmeter described previously is most satisfactory for this purpose.

Peak voltages and currents are best measured through the use of a vacuum-tube voltmeter with a large capacitance shunted by an extremely high-resistance d-c voltmeter (see Sec. 40). A cathode-ray oscillograph is also useful in such studies.

VOLTAGE-MEASURING INSTRUMENTS

29. Use of Current Meters to Indicate Voltage. All current-measuring instruments having a sensitivity in milliamperes may, with the addition of suitable series resistance,

be used to indicate potential. The current drain of the instrument must be sufficiently low to abstract negligible energy from the circuit, as otherwise corrections must be made. With modern instruments of high sensitivity this requirement can usually be met.

30. Direct measurements of voltage are obtainable through electrostatic means, but the instruments are of limited utility because of their low torque and because the minimum ranges are rarely under several hundred volts. They are essentially instruments for the research laboratory.

Table 8. Useful Technical Decibel Data

Power level, db	Power ratio to 0 db. Also power, mw, when 0 level = 1 mw	Voltage ratio to 0 db	Voltage—based on 1 mw in 600 ohms = zero level	Power level, db	Power ratio to 0 db. Also power, mw, when 0 level = 1 mw	Voltage ratio to 0 db	Voltage—based on 1 mw in 600 ohms = zero level
-10	0.1000	0.31623	0.24495	20	100.00	10.0000	7.7461
-9	0.1259	0.35481	0.27483	21	125.89	11.220	8.6912
-8	0.1585	0.39811	0.30839	22	158.49	12.589	9.7514
-7	0.1995	0.44668	0.34599	23	199.53	14.125	10.941
-6	0.2512	0.50119	0.38820	24	251.19	15.849	12.276
-5	0.3162	0.56234	0.43560	25	316.23	17.783	13.775
-4	0.3981	0.63096	0.48875	26	398.11	19.953	15.459
-3	0.5012	0.70795	0.54840	27	501.19	22.387	17.341
-2	0.6310	0.79433	0.61527	28	630.96	25.119	19.457
-1	0.7943	0.89125	0.69035	29	794.33	28.184	21.831
0	1.0000	1.00000	0.77461	30	1,000.00	31.623	24.495
+1	1.2589	1.1220	0.86912	31	1,258.9	35.481	27.487
+2	1.5849	1.2589	0.97514	32	1,584.9	39.811	30.837
+3	1.9953	1.4125	1.0941	33	1,995.3	44.668	34.600
+4	2.5119	1.5849	1.2276	34	2,511.9	50.119	38.822
+5	3.1623	1.7783	1.3775	35	3,162.3	56.234	43.560
+6	3.9811	1.9953	1.5459	36	3,981.1	63.096	48.875
+7	5.0119	2.2387	1.7341	37	5,011.9	70.795	54.840
+8	6.3096	2.5119	1.9457	38	6,309.6	79.433	61.527
+9	7.9433	2.8184	2.1831	39	7,943.3	89.125	69.035
+10	10.0000	3.1623	2.4495	40	10,000.00	100.000	77.461
+11	12.589	3.5481	2.7484	41	12,589.2	112.20	86.912
+12	15.849	3.9811	3.0837	42	15,848.9	125.89	96.698
+13	19.953	4.4668	3.4600	43	19,952.6	141.25	109.41
+14	25.119	5.0119	3.8822	44	25,118.9	158.49	122.76
+15	31.623	5.6234	4.3560	45	31,622.8	177.83	137.75
+16	39.811	6.3096	4.8875	46	39,810.7	199.53	154.59
+17	50.119	7.0795	5.4840	47	50,118.7	223.87	173.41
+18	63.096	7.9433	6.1527	48	63,095.7	251.19	194.57
+19	79.433	8.9125	6.9035	49	79,432.7	281.84	218.31

Electrostatic voltmeters depend on the attractive force which exists between two conducting plates between which a difference of potential exists. In their simplest form, the force of attraction between a stationary and a movable disk is balanced by a calibrated spring. The *Kelvin absolute electrometer* is constructed in this manner. The force of attraction is proportional to the square of the difference of potential between the plates. Such meters give the same indication on steady and alternating voltages and have neither wave form nor frequency error.

One type of construction, used in suspended-vane meters, is shown in Fig. 18. The stationary plates are sections of two concentric cylinders, into which the cylin-

dricial rotor turns. With the opposite poles of a magnet placed outside the stator plates, satisfactory damping is obtained from the currents induced in the loop. This type of construction is that used in the *Ayrton-Mather electrostatic voltmeter* built by the Cambridge Instrument Company.

Electrostatic voltmeters are very useful because of their high resistance and low power consumption at low frequencies. They cannot be used on high voltage at frequencies much above a megacycle, because of the rapid increase of the power loss in the necessary insulation. This loss increases directly as the first power of the frequency and the square of the voltage. A hard-rubber insulator with a power factor of 0.004 and capacitance of $10 \mu\text{f}$ will have, at a frequency of 10 Mc and voltage of 2.5 kv, a charging current of 1.5 amp, and a power loss of 15 watts, both of which values are excessive.

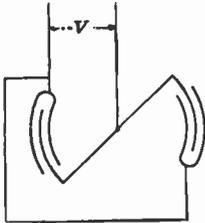


FIG. 18. Suspended-vane meter.

MEASUREMENT OF RESISTANCE, IMPEDANCE, FREQUENCY

31. Resistance. While bridge measurements of resistance give greatest accuracy direct-reading instruments are much used because there is no requirement for the manipulation of the controls, and they are widely used in production testing of resistance units as well as in general laboratory practice where the highest accuracy is not essential.

32. Ohmmeter. The simplest *direct-reading ohmmeter* consists of an ammeter and battery as shown in Fig. 19. Two readings are made, one with the terminals shorted, the other with the unknown resistance R connected. The fixed resistance S limits the current to about full-scale reading of the ammeter. The deflection is made exactly full scale by adjustment of the ammeter shunt B . The range of this type of meter is usually taken as that resistance which gives a deflection which is 5 per cent of full scale. On this basis the usual ranges are 1, 10, and 100 kilohms. Since the fixed and unknown resistances are in series, the center- or half-scale resistance value is exactly equal to the total effective resistance at the ohmmeter terminals. The two main decades of the resistance scale are symmetrically placed about the center value and together cover 82 per cent of the scale length. The voltage across the unknown resistor is not constant but increases with the resistance, being half the applied voltage at half scale.

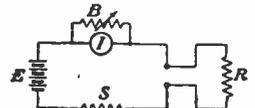


FIG. 19. Direct-reading ohmmeter circuit.

The upper limit of resistance measurements by this means depends upon the instrument sensitivity and battery voltage; a $50\text{-}\mu\text{a}$ instrument at 15 volts gives an excellent deflection on several megohms. The lower limit, since a minimum battery voltage of 1.5 volts must be used, is dependent only on the current capacity of the battery and the resistance of the leads. In general, for accurate work, the effective battery resistance must be calculated into the circuit as a part of the total series resistance.

The readings of an ohmmeter may be made independent of the applied voltage by dispensing with the controlling springs and obtaining the controlling torque from a separate coil connected across the supply voltage. Figure 20 shows the circuit used by Evershed and Vignole in their ohmmeters of this type.

This construction was first used by Evershed for an ohmmeter designed to measure high resistances up to 100 megohms. The source of voltage was a self-contained high-voltage magneto generator, giving voltages up to 500 volts. It was called a *Megger*.

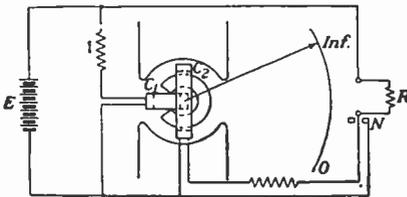


FIG. 20. Ohmmeter of Evershed and Vignole.

The same principle has now been applied to ohmmeters of lower range using battery voltages. The resistance range extends from 1 ohm to 5,000 megohms.

The upper resistance limit is greatly increased by replacing the microammeter with a vacuum-tube voltmeter (Sec. 40) connected across the fixed resistance, which may then have a maximum value of 10^9 ohms. The range will be 20,000 megohms.

The range is still further increased by using, instead of a vacuum-tube voltmeter, a degenerative d-c amplifier with a sensitivity of 0.1 volt for full-scale deflection of the indicating milliammeter. For a fixed resistance of 1,000 megohms, the current corresponding to this full-scale deflection is $100 \mu\mu\text{a}$ and 1 per cent of this, or $1 \mu\mu\text{a}$, can be detected. At 100 volts the corresponding resistance values are 1 and 100×10^{12} ohms, since at all times the fixed resistance is negligible compared to the unknown resistance. It follows then that the voltage across the unknown resistor is essentially constant. The power supply is usually tube-regulated and is provided with a protecting resistor to limit the current on a short circuit to a safe value. One decade of resistance covers 90 per cent of the scale length, so that it is necessary to provide at least two fixed resistances per decade in order to keep the fractional accuracy of reading reasonably constant.

33. Measurement of Impedance. When the voltmeter-ammeter method is used with a source of alternating voltage, the ratio of voltage to current gives the impedance of the load

$$Z = E/I \tag{2}$$

With the usual a-c instruments the corrections for the instruments are larger than for d-c instruments and more difficult to make because of their reactance. The high-resistance rectifier voltmeter and vacuum-tube voltmeter eliminate this difficulty.

The separation of impedance into its components requires the use of a wattmeter.

The connections of Fig. 21a are usually used when no correction for instrument errors is to be made, while those of Fig. 21b allow the correction to be made quite easily. For this distinction the current coil of the wattmeter is grouped with the ammeter and its potential coil with the voltmeter. As before, the impedance of the load is given by Eq. (2). Its power factor is the ratio of the wattmeter readings to the product of voltage and current.

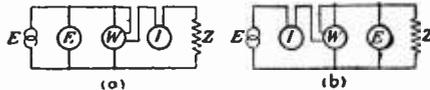


FIG. 21. Measurement of impedance.

$$\text{Pf} = \cos \theta = W/EI \tag{3}$$

where θ is the phase angle between voltage and current. The resistance of the load is

$$R = W/I^2 \tag{4}$$

and the reactance

$$X = \sqrt{W^2 - R^2} \tag{5}$$

With the knowledge as to whether the load is inductive or capacitive, its inductance or capacitance may be calculated from

$$X = \omega L = -1/\omega C \tag{6}$$

where $\omega = 2\pi f$.

34. Measurement of Capacitance. Since the power factor of the usual capacitor is small, its reactance is approximately equal to its impedance. This may be measured directly by the voltmeter-ammeter method and the capacitance calculated from Eq. (6). At a given voltage and frequency, a single ammeter reading is sufficient, and the ammeter may be calibrated to read capacitance directly.

Capacitance may also be measured on a single indicating meter whose readings are independent of the applied voltage. The moving element consists of two coils set at right angles to each other. There are no controlling springs. The connections used in the h-f Weston microfarad meter are shown in Fig. 22.

Coils C_1 and C_2 are connected across the supply voltage, one in series with a fixed capacitance S , the other in series with the unknown capacitance C . The stationary field coils F are directly connected across the line voltage. With no capacitor connected in circuit with coil C_2 , the coil C_1 sets itself in the plane of the field coils F and determines the zero of the scale.

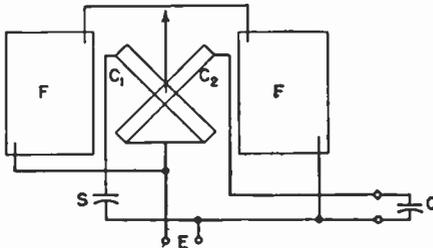


FIG. 22. High-frequency microfarad meter.

The capacitance range of the Weston microfarad meters extends from 0.05 to 10 μ f at 60 cycles, 0.001 to 0.05 μ f at 500 cycles, and 0.0005 μ f at 1,000 cycles. The applied voltage must be large enough to provide sufficient torque to give a definite reading.

35. Measurement of Power Factor. Instruments for measuring power factor are very similar to the moving-coil capacitance meters described above. The connections used in the Weston power-factor meter are shown in Fig. 23.

36. Measurement of Frequency. Frequency may be measured with an indicating instrument similar to the capacitance meter shown in Fig. 22, in which the capacitance C is fixed and the capacitance S is replaced by a resistance. The scale is, of course, calibrated in terms of frequency.

The functions of the moving and fixed coils may be transposed, the stationary part now consisting of two coils set at right angles to each other. The moving part is simply a vane of soft iron, since its sole function is to indicate the direction of the resultant magnetic field set up by the two stationary coils.

The introduction of capacitance C allows current to flow in the coil C_1 and provides an opposing torque which is proportional to the capacitance added. The resulting deflection is, of course, just as dependent on frequency as on capacitance, so that any particular instrument must be used on the exact frequency for which it was calibrated. The l-f Weston microfarad meter has the moving coils connected in series instead of in parallel with the field coils.

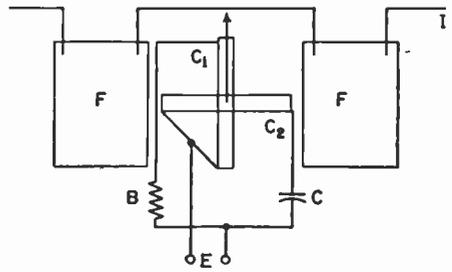


FIG. 23. Power-factor meter.

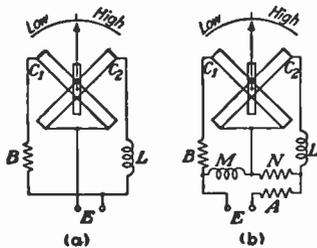


FIG. 24. Frequency meter. (Weston.)

Frequency meters¹ utilizing direct indicating instruments in conjunction with electronic circuits extend the useful range down to 3 cycles full scale and up to 60,000

¹ The following paragraphs dealing with vacuum tubes were contributed by A. G. Bousquet of the General Radio Co.

of such a frequency meter are shown in Fig. 24a. The tendency of the vane toward rotation is overcome in the Weston frequency meter by decreasing the phase difference between the currents in the two coils as shown in Fig. 24b. The rotation of the magnetic field is no longer uniform. The vane, being long and narrow, takes up a definite position, its inertia preventing it from following the irregular rotation of the magnetic field. The frequency range of the instrument is about 30 per cent of the mid-scale reading. These meters are usually built for the commercial frequencies 25 and 60 cycles. The General Electric Co. has built them for higher frequencies, up to 2,000 cycles.

cycles full scale. The electronic circuit generates pulses which are all of identical shape but which depend entirely for their frequency on the frequency of the applied voltage. The identical pulses are rectified and "counted" on a d-c instrument whose deflection is proportional to the number of pulses per second. The instrument is calibrated in terms of frequency in cycles per second.

In the circuit¹ of Fig. 25, the applied voltage is amplified and the peaks of the amplified wave are "clipped" by a diode. In successive stages, the resultant pulse shape is amplified, limited in its peak amplitude by overloaded amplifiers, and restricted in its pulse length by appropriately selected circuit constants. The pulses are then "counted" by the vacuum-tube voltmeter circuit. Each pulse charges the capacitor through one of the diodes; the capacitor discharges through the other diode, and the d-c component of the current deflects the d-c indicating instrument. Since the

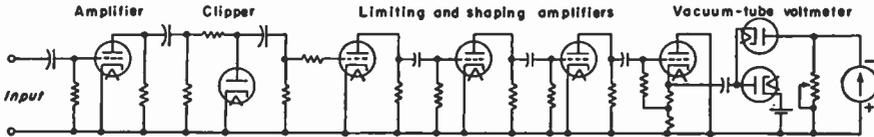


FIG. 25. One form of frequency meter.

average reading is proportional to the number of pulses per second, the instrument may be calibrated in cycles per second.

In another circuit arrangement (Fig. 26), the voltage is amplified and applied to the input of a modified Eccles-Jordan circuit which oscillates for 1 cycle each time a positive pulse (or the positive half of an alternating voltage wave) disturbs its input circuit. The frequency of oscillation is determined by the frequency of the applied voltage, but since the circuit truly oscillates, the wave shape of its output pulses is independent of the wave shape of the applied voltage. The positive pulses from the output are applied to the grid of a tube which has been biased beyond cutoff; plate current flows only while the positive pulses are applied. A d-c instrument connected in the plate circuit of the tube will be deflected each time a current pulse flows in the plate resistor, and the average instrument reading will depend on the rate at which

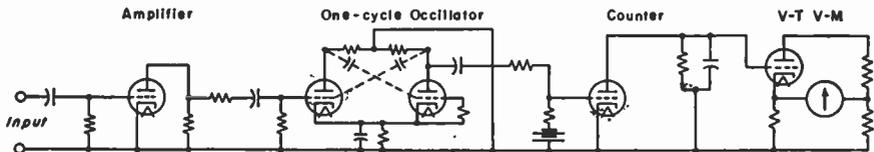


FIG. 26. Counting rate meter or frequency meter.

the current pulses flow. Therefore, the instrument may be calibrated in terms of frequency in cycles per second. The d-c instrument may be, as in the figure, a vacuum-tube voltmeter. The capacitor across the plate resistor smooths out the meter readings; this may be important at low frequencies when the d-c indicating instrument is not sufficiently damped. The circuit of Fig. 26 was developed specifically for determining the rate of emission of beta and gamma rays from radioactive substances.²

Frequency meters that make use of *vibrating reeds* are also constructed. A series of reeds, whose natural frequencies of vibration differ by regular intervals, are arranged in a line or in a circular arc in the order of ascending frequency. They are mounted on a suitably shaped electromagnet, whose winding is connected across the supply voltage of unknown frequency. That reed, having a natural frequency nearest to the supply frequency, will vibrate with an easily visible amplitude, and the frequency

¹ *Gen. Radio Experimenter*, XX (9), February, 1946.

² *Rev. Sci. Instruments*, 17 (9), 323-333, September, 1946.

intervals between adjacent reeds are sufficiently small, compared to their damping, so that at least one will always vibrate.

MOVING-DIAPHRAGM METERS

37. The telephone is a very sensitive galvanometer, in which the indication of motion is acoustic. It is essentially a moving-iron vibration galvanometer, polarized with a permanent magnet. Its construction is shown in Fig. 27. The amplitude of vibration is proportional to the product of the steady flux in the air gap produced by the permanent magnet and the alternating flux produced by the coils carrying the a.c. The latter flux is much increased by placing the coils on laminated soft-iron pole

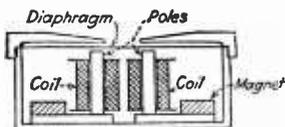


FIG. 27. Construction of telephone.

pieces. The reluctance of the hardened steel magnet to the alternating flux is so great that most of the a-c flux passes across the gap at the base of the pole pieces. This gap is made the proper length to make the product of the two fluxes at the diaphragm air gap a maximum. The diaphragm is a thin steel disk clamped at its outer edge. Its natural frequency of vibration is determined by its mass and stiffness. For silicon steel 0.01 in. in thickness, this frequency is about 900 cycles. By plugging the orifice in the carpiece, the natural frequency may be increased by as much as 50 per cent. The damping of the diaphragm is very small, being mainly due to the eddy-current losses in the iron. The variation of amplitude with frequency is a sharp resonance curve. Figure 28 shows such a curve for a Western Electric telephone. The damping is little affected by changes in stiffness and natural frequency. The impedance of a telephone winding increases with frequency in a regular way, except around the resonance frequencies. The resistance and reactance are generally of the same order of magnitude, so that its lag angle is about 45 deg. At a frequency of 1,000 cycles they are about ten times the d-c resistance of the winding. Near resonance the motion of the diaphragm introduces a counter emf into the circuit which is usually interpreted as additional resistance and reactance. These terms are referred to as *motional values*. In telephones of low damping, they may be as much as 70 per cent of the normal values. The actual numerical value of the resistance and reactance depends on the number of turns with which the magnets are wound. The d-c resistance varies from 30 to 1,000 ohms. The sensitivity of telephones is somewhat indefinite because it depends on the acuteness of hearing of the observer. It is usual to express it as the current necessary to produce a just audible response. Because of the existence of a threshold of hearing, this minimum current is reasonably definite and reproducible, at least for any one person. Values of this minimum current, together with the corresponding voltage, resistance, and power are given in Table 2 for a Western Electric receiver. It is much more sensitive than any vibration galvanometer and at its resonant frequency is not far behind a good d-c galvanometer.

38. Other Types of Telephones. It is possible to use nonmagnetic materials for the diaphragm by providing a separate steel armature so shaped and clamped that its natural frequency is higher than that of the diaphragm, to which it is attached by a stiff rod. When mica is used for the diaphragm, both the sensitivity and the selectivity are greater than for steel. On the other hand, the resonance curve can be broadened by using a corrugated diaphragm of suitable material.

The steel armature can be replaced by a coil carrying the a.c., which then may vibrate in the field of a moving-coil galvanometer. A light paper cone attached to the

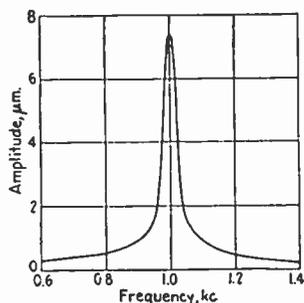


FIG. 28. Resonance curve of Western Electric telephone.

coil acts as a diaphragm. There is no single natural frequency, so that over a wide frequency range the sensitivity is essentially constant.

The piezoelectric effect exhibited in certain crystals is also used as the basis for a telephone. Rochelle salt crystals are used rather than quartz because of their greater piezoelectric effect. The construction is the same as is used in crystal microphones. The frequency characteristic of a telephone made in this manner is remarkably constant over the whole a-f range extending from 100 cycles to 5 kc. Its impedance decreases with frequency because it is essentially a capacitor with crystal dielectric. In this respect it behaves in just the opposite manner from a permanent magnet telephone.

39. Thermophones. When a fine wire is heated by the passage of a.c., sound waves are produced in the surrounding air if the heat capacity of the wire is so small that the temperature of the surface of the wire follows the cyclic variations of the current. Instruments of this sort have been constructed, using gold foil as the heater. They are called *thermophones*. Their sensitivity in terms of sound energy is low. But they can be made small enough to be placed in the ear, so that their over-all sensitivity is quite satisfactory. Their response decreases slowly as the frequency is increased. The theory of this instrument has been studied in considerable detail because of its use as a standard in the production of sound.

ELECTRON-TUBE METERS

While the direct-reading moving-coil instrument has its limitations, fortunately its range of usefulness can be vastly broadened by incorporating it in a vacuum-tube circuit. The frequency range may then be extended to as high as 300 Mc; the input resistance may be as high as 10^{16} ohms; the sensitivity as low as 10^{-16} amp.

40. D-c Vacuum-tube Meters. A simple form of vacuum-tube voltmeter is merely a triode with voltage terminals at the grid and a d-c current-indicating instrument

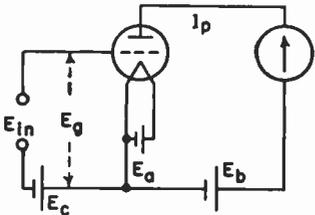


FIG. 29. Direct-current vacuum-tube voltmeter.

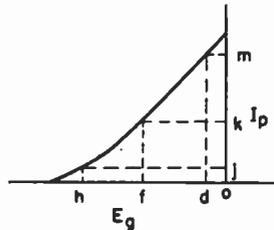


FIG. 30. Plate-current—grid-voltage characteristic.

in the plate circuit. Since the d-c plate current is a function of the d-c voltage applied at the grid, the instrument may be calibrated to indicate directly the d-c voltage applied to the grid. The connections for such a circuit are shown in Fig. 29. The vacuum tube must, of course, be powered from batteries (E_a , E_b , and E_c in the figure) or from a line-operated supply. The plate-current-grid-voltage characteristic is given in Fig. 30. The grid bias (E_o) determines the operating point on the characteristic. If a mid-point (E_f in the figure) on the characteristic is selected for the grid bias, the instrument may be calibrated for both positive and negative applied voltages.

The plate-current component of interest is the change in plate current: I_k to I_m when a positive input voltage is applied and I_k to I_j when a negative input voltage is applied. To increase the over-all sensitivity of the system, it is desirable to suppress to zero the meter reading (I_k) obtained when no voltage is applied to the input terminals, so that the entire scale may be devoted to indicating the plate-current change.

The meter reading may be suppressed mechanically so that the zero of the calibrated

scale coincides with the electrical zero, or the suppression may be obtained electrically as shown in Fig. 31.

If a resistor is common to both the plate and the grid circuits (R_k of Fig. 32) the voltage drop in this resistor due to plate current introduces an added grid bias, and the circuit is said to be degenerative. As a net result, the sensitivity of the vacuum-tube voltmeter is reduced, but the calibration is less dependent on the tube characteristics and hence on supply voltages. The degenerative resistor also provides a ready means for changing ranges.

The sensitivity of the voltmeter is directly dependent on the transconductance of the vacuum tube, where the transconductance is the ratio of plate-current increments

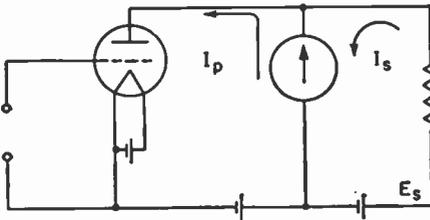


FIG. 31. Electrical suppression of plate current.

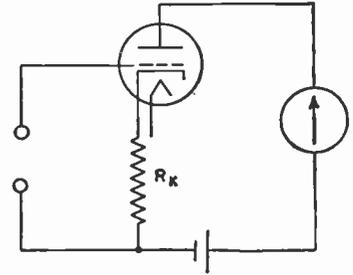


FIG. 32. Degeneration in a vacuum-tube voltmeter.

to grid-voltage increments. If the tube of Fig. 32 is replaced by a multitube amplifier, the net effective transconductance is increased enormously (10,000 fold, for example) and full-scale meter deflection of a rugged 5-ma meter may be obtained for voltages as low as 10 mv at the input terminals. With a standard resistor at the input terminals, the device becomes a microammeter and, if the current terminals are at the grid and cathode as shown in Fig. 33, the effective input resistance is reduced to a relatively low value because the voltage drop in the standard resistance R_s is balanced by the almost equal voltage drop in the degenerative resistance R_k . However the voltmeter is unduly sensitive to a small a-c voltage placed across the input terminals, because such a voltage is not degenerated but experiences the full gain of the amplifier.

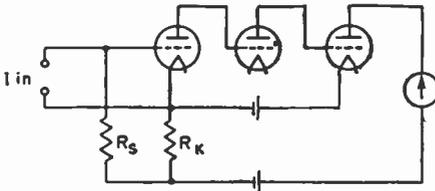


FIG. 33. Sensitive microammeter.

The major advantage of the d-c vacuum-tube voltmeter is the high input resistance. With no special precautions, the input resistance may be 100 megohms or more. In wide-range commercial voltmeters of careful design, the input resistance is of the order of 5,000 megohms. When the vacuum tube is specially designed and the circuit potentials are kept below ionizing levels, the input resistance may be as high as 10^{16} ohms. It must be remembered, however, that the final accuracy of the direct-reading vacuum-tube meter is never any better than the accuracy of its moving-coil instrument, and if proper care has not been taken in the design, the over-all accuracy may be seriously impaired.

41. A-c Triode Vacuum-tube Meters. If an a-c voltage is applied to the input terminals of the circuit shown in Fig. 29 and if the grid voltage E_g is adjusted to operate at point h of the characteristic (Fig. 34), the negative excursions of the applied

Example: For an effective over-all transconductance of 100 mhos, a value of 2 ohms at R_k , and a 1,000-megohm resistor at R_s , the full-scale sensitivity is $10 \mu\mu\text{a}$ (10^{-11} amp), with an effective input resistance of 5 megohms, instead of the 1,000 megohms that the conventional circuit would have.

The major advantage of the d-c vacuum-tube voltmeter is the high input

a-c voltage will decrease the plate current by only a small amount while the positive excursions will allow a relatively large plate current to flow. The instrument in the plate circuit will indicate the average plate-current change ($I_t - I_q$) and may be calibrated in terms of applied a-c voltage.

The operating point on the characteristic (Fig. 34) determines the type of response that will be obtained. If the bias is appreciably to the left of h (more negative than E_h), only the positive peaks of the sine wave will cause current to flow and the instrument will be a peak-reading voltmeter. If the tube is biased more positively than E_h , the response will depend to a greater extent on both halves of the cycle and will more nearly indicate the rms voltage value of both halves of the wave in spite of moderate wave-form distortion. With the bias at E_h , the resultant calibration will follow approximately a square law for small voltages because the characteristic at this point approximates a hyperbola. For large voltage swings, however, the calibration will be more nearly linear because the characteristic is nearly a straight line over the wider range.

The a-c vacuum-tube voltmeter may be operated over a wide frequency spectrum but, to avoid resonances in the plate circuit which might limit the useful range to about 100 kc, it is necessary to shunt the plate circuit with a capacitance whose impedance is low compared to the plate-circuit impedance. With the proper shunting capacitance, the calibration will be correct up to frequencies as high as 50 Mc. At the other end of the frequency spectrum, the difficulty lies in obtaining sufficient capacitance since, of course, the impedance of the capacitor increases as the frequency is decreased.

Various ranges may be obtained by introducing different values of degenerative resistance (Fig. 32). It is important to ensure that no shunt capacitance appears across the resistor, and this immediately becomes a serious limitation at high frequencies.

The major advantage of the triode type of a-c vacuum-tube voltmeter is its high input impedance. The frequency range, though great, is not so wide as the range covered by the circuits in Sec. 42. The triode circuit is not too well adapted for changing voltage ranges.

42. A-c Diode Vacuum-tube Voltmeters. If a two-electrode tube, or diode, is connected in series with a d-c instrument as in Fig. 35a, any a-c voltage applied across the series circuit will be rectified and the d-c instrument will deflect.

This most simple vacuum-tube voltmeter responds to the average of the positive half cycle of the a-c voltage. It is not a high-impedance device, and the source must provide a d-c path for the diode currents.

The voltmeter responds to the average of the positive half cycle at the low frequencies, but because of spurious shunting circuit capacitance, it responds more nearly to the peak of the wave at high frequencies. If, however, the shunting capacitance is sufficiently great, as in Fig. 35b, peak response may be maintained to quite low frequencies and the device then is useful as a peak voltmeter over a wide frequency

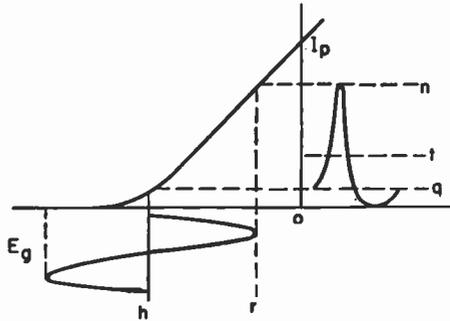


FIG. 34. Characteristic of a-c vacuum-tube voltmeter.

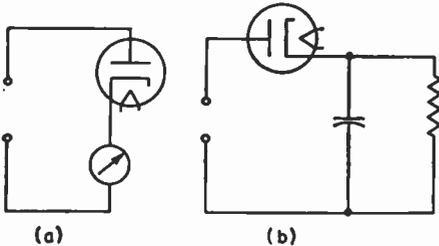


FIG. 35. Simple diode voltmeter.

range. With the diode and the capacitor interchanged as shown in Fig. 36, the a-c vacuum-tube voltmeter is a peak-reading device, which in addition provides a path for diode currents without requiring a d-c path through the source. Though satisfactory for many purposes, the input impedance is still quite low and is approximately one-half of resistance R .

If instead of measuring the d-c current through R , the d-c voltage across R is measured by a d-c vacuum-tube voltmeter, as in Fig. 36b, the value of R may be relatively very great with correspondingly high input impedance for the a-c voltmeter. During the first positive half cycle, C of Fig. 36b becomes charged to the peak value of the applied voltage. Only a small portion of the charge leaks off through R before the next positive half cycle replenishes the charge; hence the voltage appearing across the diode is the d-c voltage across C in series with the a-c source voltage.

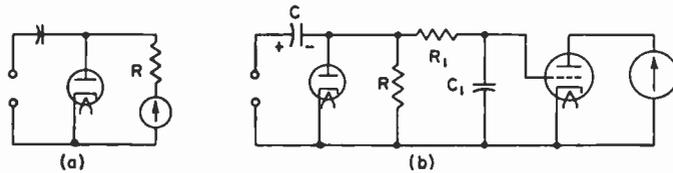


FIG. 36. Improved diode voltmeters.

Because of the RC filtering action, the a-c component is eliminated and the net voltage across C_1 at the d-c vacuum-tube voltmeter is the d-c voltage, which is equal to the positive peak of the applied a-c voltage. Commercial instruments using this basic circuit cover a frequency range from 20 cycles to 300 Mc and a voltage range of from 0.5 volt full scale to 150 volts full scale. Obviously, the upper voltage limit may be extended by means of voltage dividers, preferably of the capacitive type because of their frequency characteristics. The lower voltage limit likewise may be extended simply by preceding the vacuum-tube voltmeter by a highly stable amplifier. The amplifier design usually demands a compromise between sensitivity and frequency range.

A limitation in vacuum-tube voltmeter design is stability. A solution is balanced-circuit operation as shown in Fig. 37. Any fluctuation in the power supply affects both sides of the circuit equally with no resultant change in meter reading.

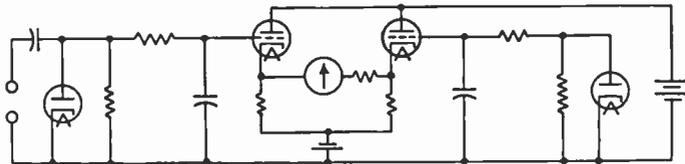


FIG. 37. Balanced vacuum-tube voltmeter.

The upper frequency limit of the diode voltmeter is determined (1) by the transit time of the electrons in their flight from the diode cathode to the plate and (2) by the series-resonance effects at the terminals. At low voltages, the two effects tend to cancel each other; at high voltages, the transit-time effect is relatively negligible.

43. Electron-stream Meters. In the cathode-ray tube, a stream of electrons impinges on a luminescent screen and forms a brilliant spot. Since an electron stream may be deflected by a magnetic or an electric field, fields due to alternating voltages will cause the spot on the screen to trace a figure which indicates the magnitude of the voltages and the phase relation between them. Some of the many varied uses include frequency intercomparisons, wave-form determinations, modulation measurements, radar applications, and television pictures.

A cathode-ray tube is shown schematically in Fig. 38. Electrons emitted from a

hot cathode are accelerated by a positive potential applied to the anode *A*. Since the anode is a cylindrical electrode disposed coaxially about the stream of electrons, many of the accelerated electrons will have sufficient velocity to continue through a small hole in the anode, finally being stopped at the screen *S*. The beam is naturally divergent because of the mutual repulsion of the individual electrons and must be focused on the screen in some manner to obtain a small sharp spot and eventually to "write" a fine line. In the earlier tubes, positive ions in the residual gas of the tube exerted a repulsive force on the electrons and prevented their divergence. The method was not too reliable and limited the upper frequency for sharp focusing to perhaps 100 kc because of the relative slowness of the ionization process.

The beam may also be focused by a longitudinal magnetic field or a radial electric field, the latter being the more convenient. For this type of focusing, the gas pressure is reduced to the minimum necessary to prevent an accumulation of negative charge on the screen. A second anode made up of two sections is mounted coaxially about the beam with one section on either side of the first anode *A*. The second anode may have a positive potential between four and five times that of the first anode, and the beam is focused by adjusting the difference in potential between the first and the second anode. In some designs the enlarged conical end of the tube is lined with a conducting layer which is connected to the second anode. The anodes may be disks or short cylinders with secondary emission shields and masking apertures to restrict the beam diameter.

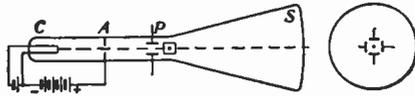


FIG. 38. Electron-stream meter or cathode-ray tube.

The cathode is usually of the oxide-coated type with a separate heater which, aside from its high efficiency in producing electrons, operates at a temperature sufficiently low so that light from it does not illuminate the screen. It is surrounded by a cylinder which serves as control electrode. A negative voltage on this electrode is used to determine the spot brilliance. It may be used to extinguish the spot completely; a timing pulse applied to the control grid permits measuring recurrence time of phenomena to a fraction of a microsecond.

The electron stream may be deflected by a transverse magnetic or electric field, applied beyond the anodes in the region where the electrons have a constant velocity. The losses inherent in the coils necessary to produce a transverse magnetic field limit their use to special cases. The transverse electric field is applied through four deflecting plates symmetrically disposed around the tube axis. When a difference of potential is applied to either pair of opposite plates, the stream of electrons is deflected toward the positive plate through an angle proportional to the strength of the electric field. The bright spot on the fluorescent screen, where the electrons strike the screen, then moves proportionally. A voltage applied between the other pair of plates produces a deflection of the spot in a direction at right angles to the first deflection. The deflection at the screen is inversely proportional to the higher anode voltage. It is of the order of 1 in. per 100 volts for an anode voltage of 1,000 volts.

When an alternating voltage is applied to a pair of plates, the electric field set up between the plates is continually varying in magnitude and direction. The stream of electrons is deflected back and forth between the plates, and the spot of light is drawn out into a line symmetrically disposed about the undeflected spot, provided the pair of plates is grounded at a point midway in potential between them. An alternating voltage applied to the other pair of plates will produce a line at right angles to the first. If the two voltages are applied to the two pairs of plates simultaneously, the electron stream follows the instantaneous resultant force exerted by both fields and traces on the screen a pattern which is closed, and therefore appears stationary, when the frequencies used bear a simple relation to one another. These patterns are called *Lissajous*¹ figures. For two equal frequencies the pattern is an ellipse of varying

¹ Barton, "Textbook on Sound," pp. 555-557, Macmillan & Co., Ltd., 1922.

eccentricity which at the extremes becomes a straight line or a circle. The exact figure is determined by the phase difference of the two voltages. For other ratios of the two frequencies the patterns become reentrant. For the general case the ratio of the number of loops formed on adjacent sides of the pattern is that of the two frequencies.

44. Timing Axis. Since the electron stream can follow accurately all variations in applied voltage, it is only necessary to spread out the line of light which it produces on the screen into a two dimensional picture to make visible its exact wave form. The second voltage of the same frequency giving the elliptical pattern just described does this but in such a manner that the whole pattern must be redrawn to be easily interpreted. The time axis, which the second voltage provides, should be linear, not sinusoidal, and its return to zero value should be instantaneous.

A very convenient circuit for this purpose employs a neon tube as shown in Fig. 39. The potential across C builds up according to an exponential law determined by the time constant CR of the circuit, which over the first part of its range is nearly linear. At some potential between 100 and 300 volts, dependent on the shape of the electrodes and the pressure of the gas, the neon tube breaks down, and the capacitor discharges very rapidly. At some lower voltage the neon tube goes out, and the charging process is resumed. If R is replaced by a diode, the curvature of the exponential law of charging may be partially compensated by the changing resistance of the tube as the voltage across it is varied. The frequency at which C charges and discharges depends on the

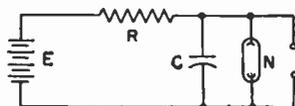


FIG. 39. Timing circuit for cathode-ray tube.

time constant CR of the charging circuit and is controlled by varying these quantities. Frequencies covering the range from 1 to 150,000 cycles are attainable. The wave form thus spread out on the screen will drift along the time axis unless the two frequencies are exactly equal or are simple multiples. It is very convenient to have the pattern stationary.

The two frequencies may be synchronized by using a thyratron in place of the two-electrode neon tube. Some voltage from the source of the wave form under observation is applied to the grid of the thyratron. When the control circuit is adjusted to produce approximately the correct frequency, this added voltage is sufficient to trigger off the discharge and maintain exact synchronism.

A time axis may also be obtained by viewing the screen on a *revolving mirror*. The pattern will be stationary when the speed of revolution of the mirror is an exact multiple of the frequency of the given wave.

Transient phenomena may be studied by photographing the single trace of the electron stream as spread out by any of the methods of obtaining a time axis just described. The time axis may also be obtained by moving the photographic film itself. In this case, and also for the revolving-mirror method, the screen must be of the type in which the fluorescence does not persist, else the trace on the film will be blurred. Screens with persistence times as short as 5 μ sec and as long as several minutes are available. The latter are useful in viewing very l-f phenomena and in radar, where it is helpful in reducing flicker.

45. Wave Analysis. When the time-axis voltage applied to one pair of plates of the cathode-ray tube is linear and its return to zero is instantaneous, the wave shape of the voltage applied to the other pair of plates is drawn out on the screen of the cathode-ray tube. If the applied voltage is purely sinusoidal, the pattern on the screen is a pure sinusoid. If the harmonic content is greater than 5 or 10 per cent, the pattern will show the distortion, and an experienced technician may estimate quite closely the nature and amount of distortion. The method facilitates amplifier and oscillator design since changes in wave shape are immediately evidenced on the screen.

Quantitative wave analysis implies a tabulation of the amount of second, third, fourth, etc., harmonic frequency present. A vacuum-tube voltmeter that may be tuned in turn to the fundamental frequency and to each harmonic frequency will

supply all the required data. There are design difficulties in providing a variable a-f filter that peaks with a flat top at the desired frequency and introduces considerable attenuation to all other frequencies. A successful solution consists in beating the unknown component of signal voltage against voltage from a local oscillator whose frequency may be adjusted to obtain a resultant beat note of predetermined fixed frequency and in measuring the voltage of the beat frequency with a vacuum-tube voltmeter tuned to the fixed frequency by means of carefully designed filters. For example, if any frequency from 0 to 15 kc is heterodyned with the output of a local oscillator whose frequency may be adjusted to the value between 50 and 35 kc necessary to produce, after conversion, a resultant "sum" frequency of 50 kc, then the vacuum-tube voltmeter need only be permanently tuned to 50 kc and the tuning problem is simplified. The tuned circuit may consist of two quartz-crystal filters tuned to either side of 50 kc, and a third quartz-crystal filter tuned exactly to 50 kc. Commercial equipment embodying these design principles is available for measuring signal components between 30 μ v and 300 volts over the a-f range.

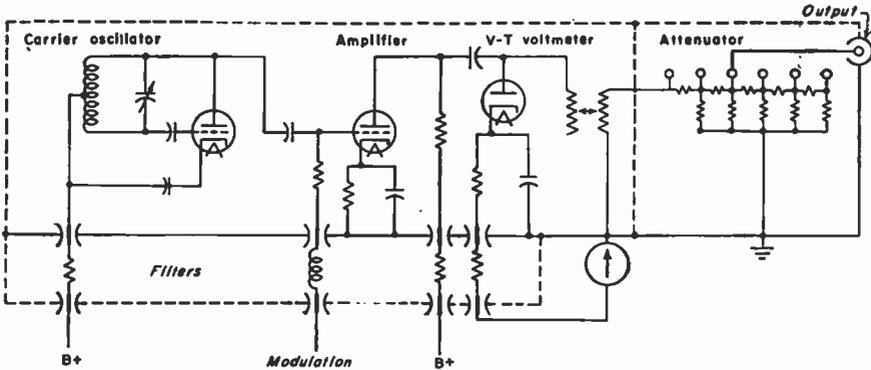


Fig. 40. Schematic diagram of typical standard-signal generator.

Some measurement problems are more concerned with the total harmonic distortion present in the signal rather than in an analysis of the individual distortion components. The measurement is readily made by filtering out the fundamental frequency voltage and measuring the remaining voltages of harmonic frequency by means of a vacuum-tube voltmeter which responds to the rms of the applied voltage. The voltage of the fundamental frequency may be eliminated by a multisection high-pass LC filter, a parallel T network (Sec. 73) embodying resistance and capacitance circuit components, or a Wien bridge (Sec. 68). The filter should transmit all pertinent harmonics equally well.

46. Standard-signal Generators. Standard methods for testing radio receiving equipment have been formulated by the IRE.¹ Most of the tests require the use of a standard-signal generator, which is essentially a miniature transmitter that may be tuned over a wide carrier frequency range, may be a-m or f-m, and is capable of delivering a signal measured to a high accuracy and adjustable over a very wide voltage range, e.g., 0.1 μ v to 2 volts. Since the output voltage must be adjustable to extremely low levels, the design must ensure a minimum of leakage from the high-level circuits in the instrument.

A schematic of a typical a-m generator is given in Fig. 40, where the carrier frequency circuits are completely enclosed in a shielded compartment which in turn is mounted with the accessory circuits (power supply, meter, modulation oscillator, etc.) in a well-shielded cabinet. All leads from the inner shielded compartment are carefully filtered to prevent leakage. The resistors of the attenuator are of low reactance

¹ "Standards on Radio Receivers," IRE, New York.

construction and are mounted in segments of a cast housing to shield the switch contacts from each other and from the switch blade. The voltage level at the input to the attenuator system is measured by means of a diode voltmeter. Since the attenuation introduced by the attenuator is calibrated, the voltage appearing at the output jack is accurately known, and if a terminated cable is used to connect the signal generator to the receiver, the voltage at the receiver is equally well known. Resistance-type attenuators may be designed to yield good results at frequencies as high as 50 Mc. At higher frequencies, capacitance attenuators or, better still, inductance attenuators are used. In the inductance attenuator, a small loop at a low impedance

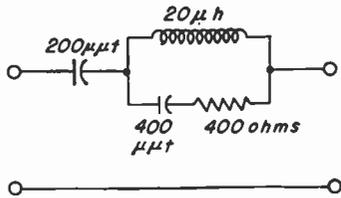


Fig. 41. Standard dummy antenna.

level is coupled to the amplifier (or oscillator) and the output voltage is varied by changing the amount of coupling. When the mechanical design of the coupling system is carefully done, the output will vary logarithmically with the linear position of the coupling coil.

The voltage from the standard-signal generator is usually coupled to the receiver under test through a standard dummy antenna, Fig. 41, whose circuit constants are selected to simulate the characteristics of the average antenna at frequencies of 500 kc and above.

The signal generator system is used mainly to determine the sensitivity and selectivity of the receiver under test at various carrier frequencies within its range.

COMPARISON MEASUREMENTS

47. Comparison of Voltages. A steady voltage may be compared with the difference of potential across a resistance carrying current by the use of the simple potentiometer shown in Fig. 42a.

A battery E_1 causes a current I to flow in R . The unknown voltage E is connected to R through a galvanometer, and R is adjusted to give no deflection of the galvanometer. The voltage E is then equal to the potential drop IR . A second voltage E' may then be made equal to a different potential drop IR' . The two currents in the two cases are the same because at balance no current flows in the galvanometer circuit. The two voltages are thus proportional to the two resistances. The potentiometer may be made direct-reading in voltage by using a standard cell for one of the comparison voltages and connecting it across such a portion of the resistance that the current must be adjusted to a predetermined decimal value in order to obtain balance. The unknown voltage is then connected through the galvanometer and balance is restored by adjustment of R , which may now be calibrated directly in volts. Connections for this type of measurement are shown in Fig. 42b.

Two alternating voltages may be compared by the potentiometer principle only when they have the same frequency and the same phase. They must at every instant be equal and opposite in order that the galvanometer in series with them shall show no deflection. Hence the potentiometer current must be taken from the same source as the voltage to be measured, and some form of phase-shifting device must be provided for which the output current is independent of its phase.

Drysdale used a two-phase induction regulator, feeding one phase through a resistance and the other through a capacitance in order to obtain the two currents in quadrature. Such a device P is shown in Fig. 43 connected to a d-c potentiometer. G_A is an a-c galvanometer having a sensitivity comparable to that of the d-c galvanom-

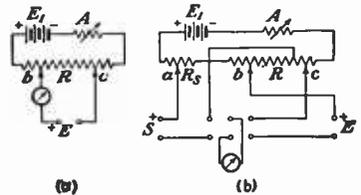


FIG. 42. Potentiometer types: (a) simple; (b) with standard cell.

eter G_D . Since there is no standard of a-c voltage, a standard cell is used to adjust the potentiometer current to its proper value. This value is read on a transfer ammeter I , which may be either of the electro-dynamometer or insulated heater thermocouple type. Its zero may be suppressed mechanically to give the effect of a longer scale and hence a greater accuracy of reading. Switches K and K_1 are then thrown to connect the potentiometer to the a-c voltages and the a-c current adjusted to produce the same deflection in ammeter I . Vacuum-tube voltmeters and rectifier voltmeters whose resistances are large compared with the resistance of the potentiometer may be calibrated directly without using the phase shifter, by connecting them directly to the terminals E . The voltage applied to them may be calculated from the settings of the contacts b and c .

48. Comparison of Frequencies. Two nearly equal frequencies may be compared by measuring in a suitable manner their difference in frequency. When the two

frequencies are in the audible range, this difference will appear as an audible beat—a waxing and waning in intensity—which may be counted if it is less than 10 beats per second. If the beats are faster than this or if the beating frequencies are above audibility, the beat must be rectified and a beat frequency produced. This beat frequency may then be measured by a suitable frequency meter.

The accuracy of the comparison depends both on the accuracy of measurement of the beat frequency and on the ratio of this frequency to the original frequencies. The beat frequency is usually kept in the audible range.

If the two frequencies to be compared are not nearly equal, so that their frequency difference is large and above audibility, audible beats may usually be obtained between some of their harmonics. For a beat frequency b between the m th harmonic of a known frequency f and the n th harmonic of an unknown frequency f' , the expression giving f' is

$$f' = \frac{mf \pm b}{n} \tag{7}$$

the sign of b being determined by considering which harmonic, mf or nf' is the larger. Sufficient harmonics are usually present in most frequency sources for the purpose of this comparison, especially when emphasized and isolated by the use of tuned circuits. They can always be produced by the use of a rectifier tube.

In the most precise measurements the known frequency is a multiple or submultiple of a standard crystal frequency, obtained from the various multivibrators driven by the standard. For less precise work a variable standard may be used. The beat frequency is then made zero. Such a variable frequency oscillator, called a *heterodyne oscillator*, will have a limited frequency range, even though provided with multiple coils. Properly chosen for range, it may be used to measure a super-audio beat frequency, such as might be obtained when comparing two very high frequencies.

Frequency is measured in terms of inductance and capacitance by means of a tuned circuit frequency meter consisting of a variable capacitance and a set of fixed inductances. The frequency range allotted to each coil determines the accuracy of setting, which ranges from 0.1 to 0.001 per cent. Resonance is indicated in a variety of ways:

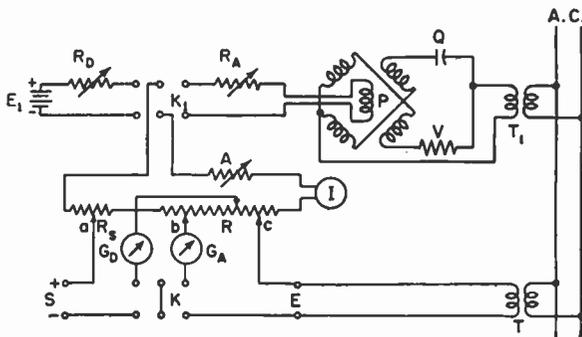


FIG. 43. Drysdale potentiometer.

thermocouple ammeter, heterodyne zero beat, or reaction on an oscillator, these being arranged in the order of their accuracy. In the third method the frequency meter is coupled closely enough to the oscillator whose frequency is being measured so that either the amplitude of its oscillations is affected or its frequency is altered. The frequency alteration is the more precise method but demands for greatest accuracy a second oscillator set at zero beat with the first. When the frequency meter is in exact resonance, the zero beat note of the two oscillators will be unaffected. In the second method a vacuum-tube oscillator is connected to the wavemeter so that it really becomes a heterodyne oscillator. A screen-grid tube, operating as a dynatron oscillator, may be connected to a frequency meter without the addition of extra coils or taps and converts it into a heterodyne-frequency meter.

49. Comparison of Impedances. An unknown resistance may be compared with a known resistance in a number of different ways. When the known resistance is variable, a *substitution method* may be employed.

The unknown resistance X is connected in series with a battery and shunted galvanometer g , the shunt resistance M having been adjusted to allow a full-scale deflection. The known variable resistance S is then substituted for X and the same

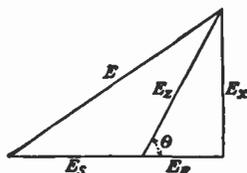


FIG. 44. Vectorial relations in three-voltmeter circuit.

current allowed to flow. Its value as thus determined is that of the resistance S . When the known resistance is not continuously variable, the value of the unknown resistance may be interpolated from the two readings of the meter. This method is frequently used for the measurement of very high resistances, such as insulation resistances from a megohm up. The known resistance is rarely larger than 1 megohm; hence under these conditions different values of the shunt M are used for the two measurements. The method is not applicable to measurements with a.c. because the phase angles of the source and load are indeterminate.

Two resistances may be compared by connecting them in series and measuring the voltage drops across them by means of a high-resistance voltmeter. Since the same current flows in both resistances, the value of the unknown resistance is

$$R = S \frac{E_R}{E_S} \quad (8)$$

where E_R and E_S are the voltages across the unknown and known resistances, respectively. Except for the case of equal resistances, the resistance of the galvanometer must be either very large compared with the resistances being measured or a correction must be made for the current taken by the galvanometer. This method may be used with a.c. to compare all kinds of impedances. Either a vacuum-tube voltmeter or a high-resistance rectifier voltmeter must be used, since correction for the current taken by the voltmeter is difficult. The polarity of the voltmeter should be maintained as in d-c measurements in order to eliminate the errors of these voltmeters due to even harmonics. The upper limit for frequency is that imposed by the frequency characteristics of the known standard and by the capacitances to ground of the voltmeter in its two positions.

The power factor of an unknown impedance may be determined by the *three-voltmeter method*, in which the voltages across the unknown and known impedances and that applied to the two in series are read. The same precautions concerning polarity and capacitances to ground apply as in the two-voltmeter method. The vectorial relations between the three voltmeter readings together with the voltage components of the unknown impedance are shown in Fig. 44.

The expressions giving the unknown impedance Z , its resistance R , reactance X , and power factor $\cos \theta$ are

$$\begin{aligned}
 Z &= S \frac{E_z}{E_s} \\
 R &= S \frac{E^2 - E_z^2 - E_s^2}{2E_s^2} \\
 X &= \sqrt{Z^2 - R^2} \\
 \cos \theta &= \frac{R}{Z} = \frac{E^2 - E_z^2 - E_s^2}{2E_z E_s} \tag{9}
 \end{aligned}$$

50. Resistance-variation Method. The total resistance of a circuit may be measured by the *resistance-variation method*. Since with a constant applied voltage the current flowing in the circuit is inversely proportional to the total resistance, the circuit resistance is given by

$$R = S \frac{I'}{I - I'} \tag{10}$$

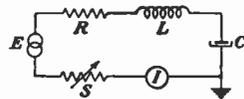


FIG. 45. Resistance-variation method.

where I is the initial current and I' the current which flows when the resistance S is added. A plot of the reciprocal of the current flowing for different values of the added resistance against that resistance gives a straight line whose negative intercept on the resistance axis is the circuit resistance. The added resistance necessary to halve the current is also the circuit resistance. This method is sometimes used to measure the resistance of a sensitive galvanometer.

The resistance-variation method may be used with a.c. provided the circuit is tuned to resonance. The necessary connections are shown in Fig. 45. By reducing the reactance of the circuit to zero, the same equations and procedure may be used as for d.c. The ammeter used is usually of the thermocouple type. Halving the current on such a meter quarters the deflection; hence this type of measurement is sometimes called the *quarter-deflection method*. The ammeter may be replaced by a vacuum-tube voltmeter connected across the capacitor. This arrangement is much more sensitive than the thermocouple ammeter and simplifies the grounding of the circuit by eliminating one series element. The upper limit for frequency is set by the frequency characteristic of the known resistance and the capacitances to ground of the different parts of the circuit. The former condition usually dominates unless the added resistance is small. The frequency limit is about 2 Mc.

Two reactances may be compared in a tuned circuit by a substitution method. The circuit is tuned to resonance both when the unknown reactance is connected in circuit and when it is disconnected. The change in reactance of the variable standard, with which the circuit is tuned, is equal to the unknown reactance. When the unknown and known reactances are both inductive or both capacitive, the value of the unknown inductance or capacitance is obtained directly, independent of frequency, the two reactances being connected in series if inductive, and in parallel if capacitive. For these pairs of measurements it is unnecessary that the currents be kept of the same value.

Air capacitors are much better standards at high frequencies than inductors; therefore, it is usual to measure an unknown inductance in terms of a variable capacitor. Small inductances are connected in series with the capacitor and large inductances in parallel with the capacitor.

The resistance of the unknown reactance may be determined by noting the current at resonance when it is connected in circuit and then by adjusting the current to this same value by adding sufficient resistance when it is disconnected. This added resistance, corrected for the change in resistance of the standard reactance with setting, is the resistance of the unknown reactance. The resistance of the standard reactors must be measured by one of the variation methods or by one of the bridge methods.

The resistance of a variable air capacitor follows a definite law, and this fact may be used in this type of resistance measurements.¹

51. Reactance-variation Method. The total resistance of the tuned circuit may also be measured by detuning the circuit. This method is called the *reactance-variation method*.² The change in reactance necessary to halve the squared current (deflection of a thermocouple meter) or to reduce the reading of a vacuum-tube voltmeter in the ratio of 1 to $\sqrt{2}$ (0.707) is equal to the resistance of the circuit. The resistance of an unknown reactance may be found by again measuring the total resistance of the circuit when the unknown is added. The difference in circuit resistance with the unknown in and out is the unknown resistance. The circuit resistance for the one case can also be found from the other by multiplying the known circuit resistance by the ratio of the voltmeter readings at resonance.

52. Susceptance-variation Method. The errors caused by capacitance to ground of the different parts of the circuit can be eliminated by connecting all parts—inductor, capacitor, and vacuum-tube voltmeter—in parallel, with one side of each grounded,

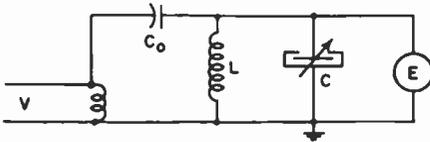


FIG. 46. Susceptance-variation method.

as shown in Fig. 46. This method is called the *susceptance-variation method*. The circuit is coupled very loosely to the generator, usually capacitively, although inductive or resistive coupling can be used. The fractional change in tuning capacitance necessary to reduce the reading of the vacuum-tube voltmeter

in the ratio of 1 to $\sqrt{2}$ (0.707) is equal to the dissipation factor ($D = R/X = G/B$) of the circuit. To increase the accuracy of this measurement, it is usual to detune on both sides of resonance and to take one-half the total change. The top of even a sharp resonance curve is broad compared to its steep sides. The dissipation factor of an unknown reactance may be found by again measuring the total dissipation factor of the circuit when the unknown is added. The difference in dissipation factor with the unknown in and out is the dissipation factor of the unknown reactance. This dissipation factor can also be found by detuning the circuit only once and noting both resonant voltages. The dissipation factor of the unknown reactance is equal to that of the circuit measured times the ratio of the change in resonant voltage to the resonant voltage of the circuit when it was not detuned. The upper limit of frequency is set by the series resistance and inductance of the standard capacitor.¹ This limit depends on the change in capacitance of the standard capacitor with the unknown connected and disconnected. It is about 10 Mc for 1,000 $\mu\mu\text{f}$ and 30 Mc for 100 $\mu\mu\text{f}$, when the standard capacitor has a series inductance of 0.006 μh .

In the case of specimens of solid dielectrics in the form of thin disks 2 in. or less in diameter, the effect of series inductance can be entirely eliminated by the use of micrometer electrodes, as described by Hartshorn and Ward.³ The circuit is first resonated with the specimen clamped between the electrodes. With the specimen removed, the movable electrode is advanced until the circuit is brought back to resonance. The dissipation factor of the specimen is calculated as previously described. The capacitance is not exactly the change in the tuning capacitance because of the finite thickness of the foil or silver-paint electrodes usually applied to the specimen. The average thickness of the specimen is measured with micrometer calipers before applying the electrodes and the air capacitance calculated. The true capacitance of the specimen is the difference in the calibration capacitances of the micrometer electrodes at their setting when the specimen was removed and at a setting equal to the thickness of the specimen, increased by the calculated air capacitance. The upper

¹ See Sec. 62.

² Sinclair, D. B., *Proc. IRE*, **26**, No. 12, 1466-1497.

³ Hartshorn, L., and W. H. Ward, *Proc. IEE (London)*, **79**, 597-609.

limit of frequency is that for which the inductor becomes a straight wire connected across the terminals of the standard capacitor. This is about 150 Mc for a capacitance range of 50 μmf . For frequencies above 50 Mc radial resistance of the electrodes introduces an error in the dissipation factor measurements. This occurs because contact to the electrodes may be at only one or a few small areas, since the sample or the electrodes are neither plane nor parallel and the current is forced to flow on the surface of all conductors because skin effect is complete. The best solution of this difficulty is to grind all surfaces plane and parallel and use no added electrode material.

53. Resonance-rise Method. When a voltage is introduced in series into a tuned circuit, the ratio of this voltage to that appearing across the tuning capacitor is approximately the dissipation factor of the circuit. This voltage is usually introduced as the voltage drop across a small resistance, as shown in Fig. 47. A commonly used value for this resistance is 0.04 ohm. The current through this resistor is measured by a thermocouple ammeter. When this current is held at a definite value, the vacuum-tube voltmeter can be calibrated in dissipation factor D or its reciprocal, storage factor Q . The latter is proportional to the voltage indicated by the voltmeter and hence has a more nearly linear scale. Since both the resistance of the coupling resistor and the standard capacitor are negligible compared to that of the inductor, the indicated dissipation factor is approximately that of the inductor. Its inductance can be calculated from the tuning capacitance and the frequency. An unknown reactance can be measured by connecting it in parallel with the standard capacitor exactly as described in Sec. 52 for the susceptance-variation method.

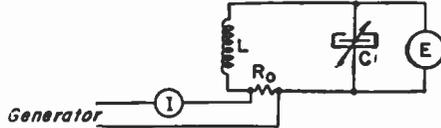


FIG. 47. Resonance-rise method.

D-C BRIDGE MEASUREMENTS

54. Whenever two resistances are compared by matching or comparing the deflections of any deflecting instrument, the accuracy of the measurement is determined by the accuracy of reading of the deflections themselves. This accuracy may be greatly increased by adopting a null method, in which a certain relation of the resistances being compared is indicated by a zero deflection. As this condition is approached, the sensitivity of the indicating instrument may be increased.

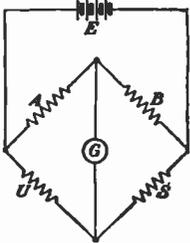


FIG. 48. Wheatstone bridge.

55. Four-resistance Network. The simple four-resistance network invented by Christie in 1833 and exploited by Wheatstone 10 years later is shown in Fig. 48.

Two paths are provided for the current, one through the ratio arms A and B , the other through the unknown and known resistances U and S . The galvanometer G is connected between the junctions of these pairs of resistances. The condition for a null deflection of the galvanometer is that these two junctions are at the same potential. Equating the voltage drops

$$AI_A = UI_U \quad \text{and} \quad BI_B = SI_S \tag{11}$$

or, since no current flows in the galvanometer,

$$\frac{A}{B} = \frac{U}{S} \quad \text{or} \quad U = \frac{A}{B} S \tag{12}$$

The ratio arms are usually only variable in steps of 10 so that the bridge is balanced by varying the known resistance S .

In commercial bridges the accuracy ranges from 0.1 to 0.02 per cent. Switching is accomplished by sliding-contact decade switches or taper plugs, and the ratio arms are

reversible. There are four to six decades in the known resistance, hundredths to hundred thousands, and up to nine ratios, 0.0001 to 10,000. Comparisons of resistances on the best bridges using sealed standards, flat mercury contacts, and a temperature-controlled oil bath may be made to 1 part in 1,000,000, which is beyond the accuracy with which the primary standard of resistance is known.

56. The sensitivity of the null detector necessary to attain a given accuracy of bridge balance is determined by the relative magnitude of the resistances of the bridge arms and the voltage applied to the bridge. The ratio of the output voltage e to the input voltage E is given by

$$\frac{e}{E} = \frac{G/B}{1 + \frac{A}{B}} \cdot \frac{A/B}{\frac{A}{B} \left(1 + \frac{S}{B}\right) + \frac{G}{B} \left(1 + \frac{A}{B}\right)} d \quad (13)$$

where G is the resistance of the null detector and d is the fractional accuracy of balance demanded. When all four arms are equal

$$\frac{e}{E} = \frac{1}{4} \frac{G/B}{1 + \frac{A}{B}} d \quad (14)$$

This ratio lies between $\frac{1}{6}d$ and $\frac{1}{4}d$ for ratios of detector and bridge-arm resistances between one and infinity. In general, its value decreases rapidly when the bridge arms are made unequal and when the detector resistance is low compared to them.

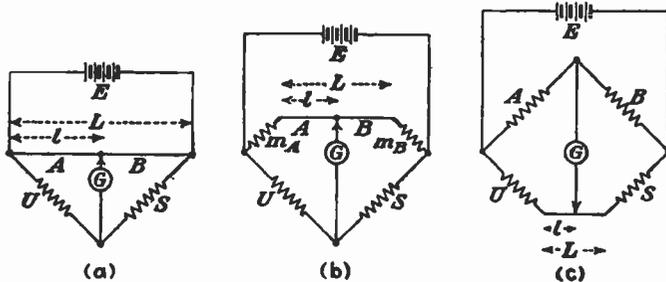


FIG. 49. (a) Slide-wire bridge; (b) bridge with extension arms; (c) Carey Foster bridge.

On this account, resistances above 1 megohm cannot be accurately measured when a d-c galvanometer is used as a null detector.

For a very high resistance detector, Eq. (13) becomes

$$\frac{e}{E} = \frac{A/B}{\left(1 + \frac{A}{B}\right)^2} d \quad (15)$$

which is independent of the ratio S/B . This condition may be realized by the use of a vacuum-tube voltmeter as described in Sec. 40. Thus for greatest sensitivity the detector should be connected from the junction of the highest resistances to the junction of the lowest. The battery, on the other hand, should be connected across the higher and lower resistance pairs, so that the amount of power drawn by the bridge is a maximum.

57. Slide-wire Bridges. When the known resistance is fixed, the bridge must be balanced by varying one or both of the ratio arms. In the slide-wire bridge shown in Fig. 49a the ratio arms A and B are parts of a single uniform resistance along which the contact of the lead from galvanometer may slide. The position of the contact

is read as a distance measured from one end, the whole length of the scale being L divisions. The value of the unknown resistance in terms of these distances is

$$U = \frac{l}{L-l} S \tag{16}$$

When the known and unknown resistances are nearly equal, the accuracy of measurement may be increased by placing extension coils in series with the slide wire as shown in Fig. 49b. The slide wire may be calibrated to read directly the percentage error of the unknown resistance U in terms of the standard resistance S .

Two nearly equal resistances may also be compared by means of the Carey Foster bridge shown in Fig. 49c. This is a slide-wire bridge in which the slide wire is placed between the two resistances being compared. Two settings of the slide wire l and l' are made with the resistances U and S as shown in Fig. 49c and transposed.

The value of the unknown resistance is

$$U = S - (l - l')\rho \tag{17}$$

where ρ is the resistance per unit length of the slide wire.

58. Kelvin Bridge. In the measurement of 0.1 ohm or less, the variation in contact resistance at its terminals and the consequent variation in the lines of current flow near the terminals may produce appreciable errors. To overcome this difficulty, low-resistance standards are always built as four-terminal resistances. All ammeter shunts are so constructed. The two potential terminals are placed between the current terminals and the resistance proper. The value of the resistance is that between the potential terminals.

Such four-terminal resistances cannot be compared on the ordinary Wheatstone bridge. They may be measured on the Kelvin double bridge shown in Fig. 50. The two four-terminal resistors U and S are connected in series, leaving an unknown resistance M between their adjacent potential terminals. The bridge is balanced by adjustment of the standard resistance S .

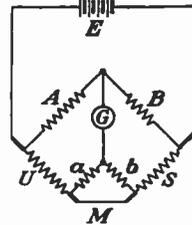


FIG. 50. Kelvin double bridge.

The value of the unknown resistance U is given by

$$U = \frac{A}{B} S \tag{18}$$

when the double ratio arms are proportional, satisfying the condition $A/B = \alpha/\beta$.

A-C BRIDGE MEASUREMENTS

59. Four-impedance Network. When an alternating voltage is applied to the simple Wheatstone bridge of Fig. 48, the conditions for balance of the bridge involve the impedances of the four arms, as shown in Fig. 51.

For a null deflection of the a-c galvanometer or telephones the two junctions, across which it is connected, must be at the same potential at all instants of the a-c cycle. Equating the voltage drops along the two parallel paths offered to the flow of the a.c.

$$Z_A I_A = Z_U I_U \quad \text{and} \quad Z_B I_B = Z_S I_B \tag{19}$$

where Z_A, Z_B , etc., replace A, B , etc., in Fig. 48.

The four impedances are vectors of the form

$$Z = R + jX \tag{20}$$

Hence, since no current flows in the galvanometer,

$$\frac{Z_A}{Z_B} = \frac{Z_U}{Z_S} \tag{21}$$

Expanding these vectors into their rectangular components the two conditions of balance are

$$\frac{R_A}{R_B} = \frac{R_U}{R_S} + \frac{X_A X_S - X_B X_U}{R_B R_S} = \frac{X_U}{X_S} + \frac{R_U X_B - R_S X_A}{R_B X_S} \quad (22)$$

If the ratio arms have no reactance, so that $X_A = X_B = 0$, these conditions reduce to

$$\frac{R_A}{R_B} = \frac{R_U}{R_S} = \frac{X_U}{X_S} \quad (23)$$

The two reactances must have the same ratio as their resistances and as the ratio arms. Considering the reactances as both inductive or both capacitive, Eq. (23) becomes

$$\frac{R_A}{R_B} = \frac{R_U}{R_S} = \frac{L_U}{L_S} \quad \text{and} \quad \frac{R_A}{R_B} = \frac{R_U}{R_S} = \frac{C_U}{C_S} \quad (24)$$

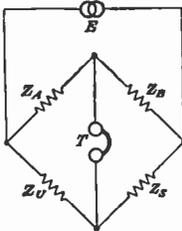


FIG. 51. A-c bridge.

respectively. These equations cover all the types of bridge measurements in which similar impedances are compared.

60. Power Supply and Null Detector. The power source at audio and radio frequencies is usually a vacuum-tube oscillator, capable of supplying several hundred milliwatts of power at varying potentials up to 100 volts. At the low audio frequencies, a-c generators with rotating parts may be used, as well as the commercial power supplies at 60 and 25 cycles. The null detector most frequently used in the a-f range from 400 to 5,000 cycles is the head telephone. Vibration galvanometers and a-c moving-coil galvanometers are used at power frequencies. Rectifier voltmeters are used for all frequencies up to 50 kc, cathode-ray and "magic-eye" tubes up to 1 Mc, and vacuum-tube voltmeters at all frequencies. At super-audio frequencies a heterodyne oscillator and detector may be used to produce an a-f beat note, which can then be observed by any of the methods described. Radio-frequency oscillators may be modulated at an a.f., usually 1 kc, and the bridge output observed on a radio receiver. All-wave receivers cover the frequency range from 10 kc to 60 Mc.

Vacuum-tube amplifiers are used with all types of null detectors to give increased sensitivity. The amount of amplification necessary to give any desired accuracy of balance may be determined by Eq. (15) when the generator is connected across resistive ratio arms. If the generator is placed across unlike arms, one resistive and one reactive, this expression becomes

$$\frac{e}{E} = \frac{Z_A/Z_B}{1 + (Z_A/Z_B)^2} d \quad (25)$$

At the most Eqs. (15) and (25) differ by only a factor of 2. These two equations hold exactly for the larger component of impedance, provided that the square of the ratio of the small to the large component is negligible compared to unity. The value of e/E for the smaller component is then less than that for the larger component by their ratio. The vibration and a-c moving-coil galvanometers are about equally sensitive, with a minimum detectable voltage of 20 μv , although a moving-coil galvanometer can be built with a sensitivity of 0.1 μv . Head telephones come next with a minimum detectable voltage of 400 μv . Then in turn come "magic-eye" tubes at 20 mv, vacuum tube and rectifier voltmeters at 100 mv, and cathode-ray tubes at 1 volt.

A considerable amount of selectivity is desirable in a null detector to eliminate the effect of harmonics in the generator, harmonics produced by nonlinearity of the unknown impedance and voltages induced in the exposed parts of the bridge circuit by external electrostatic and electromagnetic fields, usually at the commercial power

frequency. This can be provided by a tuned circuit in the amplifier or by the degenerative feedback amplifier described by Scott.¹ This latter amplifier is particularly valuable because it can be made continuously adjustable over the entire a-f range. The former gives a discrimination of at least 30 db against the second harmonic, and the latter about 40 db. The vibration galvanometer is extremely selective and offers about 70 db against the second harmonic. The a-c galvanometer is phase sensitive and responds only to that component of the unbalance voltage which is in phase with its field. It can, therefore, be made to respond to only one component of bridge balance at a time by connecting its field to a suitable phase-shifting network. The cathode-ray tube can be used in a somewhat similar manner by applying the bridge voltage to its horizontal deflecting plates as a sweep circuit through a phase shifting network.² The general pattern appearing on the screen is a tilted ellipse, which at balance reduces to a horizontal line. The phase of the sweep voltage can be so adjusted that one component of bridge balance opens the ellipse while the other tilts it.

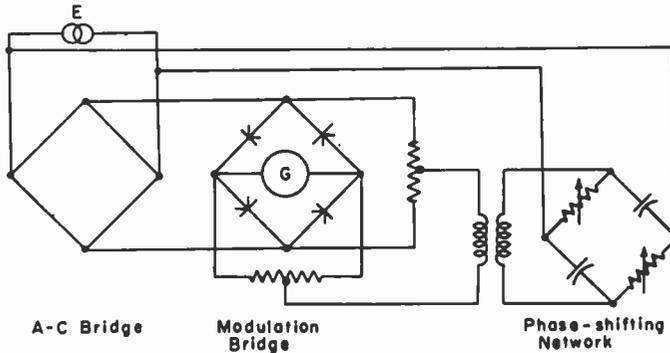


FIG. 52. Selective directional null detector.

The use of a modulation bridge with a d-c galvanometer offers a means of obtaining both a complete separation of the two bridge balances and the directional effect inherent in the d-c galvanometer when used on a d-c bridge.³ The modulation bridge consists of four rectifiers, all directed in the same way, fed both by the bridge output and by a phase-shifting network connected to the bridge supply, with the d-c galvanometer connected across opposite corners, as shown in Fig. 52. The deflection of the d-c galvanometer is proportional to the vector product of the outputs of the a-c bridge and the phase-shifting network and hence is unaffected by even harmonics and off-frequency voltages. The phase-shifting network is usually provided with two settings 90 deg apart, adjusted to match the phases of the two bridge balances. These two phases depend on the dissipation factor of the impedances being compared on the a-c bridge. The phase-shifting network must, therefore, be reset whenever there are considerable variations in this dissipation factor.

61. Bridge Transformers. Transformers are used to match the impedance of a bridge to that of the generator or detector and to isolate the bridge electrostatically. One junction point of the bridge, usually that between the two impedances being compared, is grounded, except when direct impedances are measured.⁴ The capacitances to ground of the transformer, generator, or detector not connected to this grounded junction are placed across the two bridge arms whose junction point is grounded. The effect of the ground capacitances of the generator or detector connected to the transformer may be removed by placing a grounded shield between the

¹ Scott, H. H., *Proc. IRE*, 26 (2), 226-235.
² Lamson, H. W., *Rev. Sci. Instruments*, 9 (9), 272-275.
³ Lamson, H. W., *Proc. IRE*, 28 (12), 541-548.
⁴ See Sec. 64.

primary and secondary windings. An impedance bridge with such a transformer connected across its ratio arms is shown in Fig. 53. Terminal capacitances C_{TU} and C_{TS} are placed across the bridge arms U and S . They are usually of the order of several hundred micromicrofarads and may, therefore, introduce serious errors. The direct capacitance between the two windings may be reduced to a few tenths of $1 \mu\mu\text{f}$.

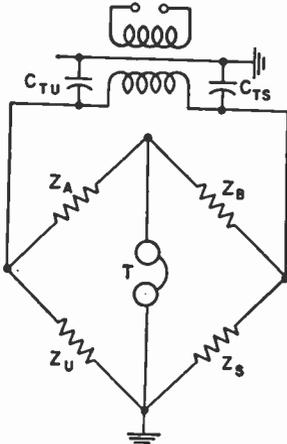


FIG. 53. Impedance bridge with transformer.

The effect of C_{TU} and C_{TS} may be reduced or either one made zero by the addition of a second shield. The two shields are symmetrically placed around the two windings, as shown in Fig. 54. C_n between the two shields may be made much smaller than the terminal capacitances and is in series with them. The

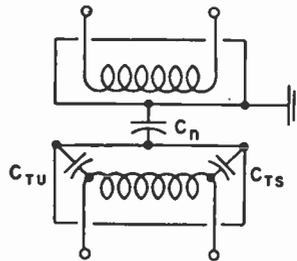


FIG. 54. Bridge transformer capacitances.

resultant terminal capacitances may be placed across either bridge arm U or S by connecting the shield around the secondary winding to one terminal of that winding. The effect of the terminal capacitances may be removed entirely from the arms U and S and placed across the ratio arms A and B by introducing a third shield between the two winding shields and connecting it to the junction of the ratio arms.

Unless the shielding of a bridge transformer is very well carried out, the direct capacitance between the two windings will still exist, reduced perhaps to a few hundredths of $1 \mu\mu\text{f}$. Even this small value will produce appreciable error in the dissipation factor balance of the bridge if this capacitance has a conductive component and if the generator is grounded, as shown in Fig. 55. C_d and its parallel resistance R_d , usually representing its dielectric loss, are placed in series with the generator voltage E across the bridge arm Z_U . It has been shown¹ that this is equivalent to placing a passive admittance proportional to C_d and R_d directly across Z_U . The effect on the reactance of Z_U is usually negligible, but not the effect on its resistance. The error can be reduced both by providing the transformer with a high step-up ratio and by using in the transformer only insulation, such as polystyrene, having a very small dielectric loss.

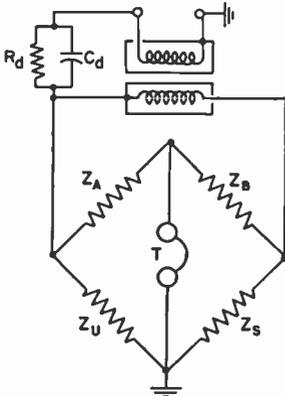


FIG. 55. Direct capacitance between windings.

generator and detector cause the more serious errors in bridge measurements. These residual reactances may be inductances in series with the bridge arms and capacitances in parallel with them. The effect of such residuals in the ratio arms may be seen by rewriting Eq. (22) of Sec. 59 in the approximate form

62. Bridge Errors. Reactances introduced into the arms of a bridge by the wiring of the bridge and by the

¹ Easton, I. G., *Gen. Radio Experimenter*, 16 (5), 4-6.

$$\left. \begin{aligned} \frac{R_A}{R_B} &= \frac{R_U}{R_S} \left[1 + (Q_A - Q_B) \frac{1}{D_U} \right] = \frac{X_U}{X_S} [1 - (Q_A - Q_B) D_U] \\ D_S - D_U &= Q_A - Q_B \end{aligned} \right\} \quad (26)$$

where the *storage factors* Q_A and Q_B and the *dissipation factors* D_U and D_S are of the form

$$Q = X/R \quad \text{and} \quad D = R/X \quad (27)$$

The errors introduced are proportional to the difference of the storage factors of the ratio arms, multiplied by the dissipation factor of the impedance arms for the reactance component, and divided by that dissipation factor for the resistance component. For impedances with small dissipation factors the error is confined to the resistance component; for impedances with large dissipation factors to the reactance component.

Residual reactances in the impedance arms produce at low frequencies errors proportional to their ratio with similar reactances in these arms. Series inductance introduces large errors in measurements of small inductances; parallel capacitance in measurements of small capacitances.

The effect of residual reactances increases with frequency, the storage factor of the ratio arms being of the form $Q = \omega L/R$ for series inductance and $Q = R\omega C$ for parallel capacitance. Hence bridges designed for operation at frequencies much above 100 kc must have equal ratio arms, because of the difficulty of equalizing their storage factors. When residual inductance in the impedance arms is in series with a capacitance, the effective capacitance of the combination is

$$C = \frac{C}{1 - \omega^2 LC} \quad (28)$$

which increases indefinitely as the resonant frequency is approached. For an inductance of $1 \mu\text{h}$, the approximate value for a constant-inductance three-dial decade resistor, and a capacitance of $1,000 \mu\mu\text{f}$ the resonant frequency is 5 Mc. Even the lowest inductance which a commercial $1,000\text{-}\mu\mu\text{f}$ capacitor has, $0.006 \mu\text{h}$, gives a resonant frequency of 65 Mc.

The errors introduced into bridge measurements by reactances in the ratio arms may be minimized by the use of substitution methods. The effect of capacitances to ground and the effect of the reactance of the leads to the known and unknown reactances will also be greatly reduced. Both reactances are connected in the same arm of the bridge, a similar reactance being placed in the other arm. Two bridge balances are obtained, one with the unknown reactance in circuit, the second with it disconnected and its impedance replaced by the known variable reactance and the added resistance. Inductances are connected in series, placing them far enough apart to reduce their mutual inductance to a negligible amount, and the unknown is removed by shorting. Capacitances are connected in parallel, and the unknown is removed by disconnecting its high-potential terminal. Both capacitors must be completely shielded and their grounded terminals connected together.

Distinguishing the values for the second balance, when the unknown reactance has been removed, by primes, the values of the unknown reactances are given by the change in reactance of the variable standards.

$$\begin{aligned} L_U &= L_S' - L_S & C_U &= C_S' - C_S \\ &= \Delta L_S & &= \Delta C_S \end{aligned} \quad (29)$$

The corresponding expressions for the resistances are

$$\begin{aligned} R_U &= R_S' - R_S + R' - R & R_U &= (R' - R) \left(\frac{C_S'}{C_U'} \right)^2 \\ &= \Delta R_S + \Delta R & &= \Delta R \left(\frac{C_S'}{C_U'} \right)^2 \end{aligned} \quad (30)$$

The squared terms appearing in the expression for the resistance of the capacitor result from the law by which the series resistance of capacitors connected in parallel is found

$$R = \frac{R_1 C_1^2 + R_2 C_2^2 + \dots}{(C_1 + C_2 + \dots)^2} = \frac{\sum_1^n R_m C_m^2}{\left(\sum_1^n C_m\right)^2} \quad (31)$$

The terms containing the resistance of the standard capacitor have disappeared because the quantity RC^2 for an air capacitor is a constant, independent of the setting of the capacitor. This follows from the more general law that for an air capacitor, in which the losses occurring in the solid dielectric are independent of the setting of the plates and for which the dissipation factor of the solid dielectric is independent of frequency, the quantity $R\omega C^2 = DC$ is constant. This law holds with increasing frequency until the losses due to skin effect in the plates and supports become appreciable.

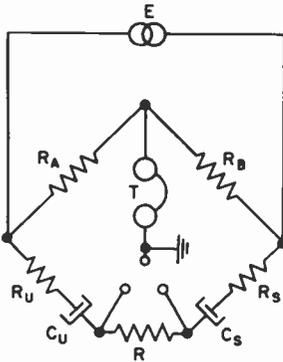


FIG. 56. Series-resistance bridge.

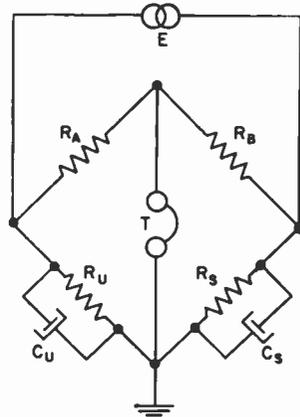


FIG. 57. Parallel-resistance bridge.

The series resistance of the plates and supports of a well-designed air capacitor is of the order of 0.02 ohm at a frequency of 1 Mc.¹ This resistance varies as the square root of the frequency because even at 1 Mc the skin effect is complete. By shortening the leads and by connecting to the stator and rotor at several points, this series resistance can be reduced to 0.005 ohm at 1 Mc.

63. Resistance Balance. When two impedances are compared on a four-impedance bridge, the conditions of balance [Eq. (24) of Sec. 57] demand that their dissipation factors be equal. Since this will not in general be the case, means must be provided for attaining the resistance balance. The simplest method is that of adding a resistance in series with that impedance having the lower dissipation factor. The connections for a capacitance bridge with the added resistance so arranged that it may be placed in either impedance arm is shown in Fig. 56. This method gives the series resistance and reactance of the unknown impedance and can be used for dissipation factors less than unity. Neither of the impedances, although essentially at ground potential, can be grounded, unless it is known that one of them, usually the standard, will always have the smaller resistance.

Added resistances may be placed in parallel with the two impedance arms as shown in Fig. 57. This method gives the parallel resistance and reactance of the unknown

¹ Field, R. F., and D. B. Sinclair, *Proc. IRE*, **24** (2), 255-274.

impedance and is best adapted to the measurement of impedances having dissipation factors greater than unity. For small dissipation factors the shunting effect of the parallel resistances is such as to reduce markedly the sensitivity of the bridge balance. One terminal of each impedance is grounded.

The resistance balance may also be made by adding suitable reactances to the ratio arms. Rosa in 1907 suggested the use of series inductance, while Thomas in 1914 used parallel capacitance, as shown in Fig. 58. The balance equations are

$$C_U = \frac{R_B}{R_A} C_S \text{ (approx)} \quad \text{and} \quad R_U = \frac{R_A}{R_B} R_S + R_A \left(\frac{C_B}{C_S} - \frac{C_A}{C_U} \right) \quad (32)$$

whence

$$D_U = D_S + Q_A - Q_B$$

The capacitance measured in this manner is the series capacitance. The parallel capacitance and resistance may be calculated from the expressions

$$C_p = \frac{C_S}{1 + D^2} \quad \text{and} \quad R_p = R_S(1 + Q^2) \quad (33)$$

Schering in 1920 used a parallel capacitance across one ratio arm in a high-voltage bridge connected as shown in Fig. 59. The generator was connected from the junction of the resistance arms to the junction of the capacitance arms, both to minimize the power losses in the ratio arms and to keep constant the voltage applied to the unknown capacitor. The junction of the resistance arms was grounded in order to keep the ratio arms and the detector at a low voltage with respect to ground.

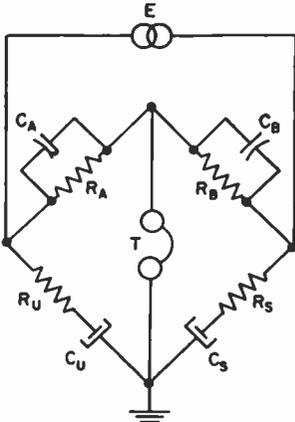


FIG. 58. Thomas's method.

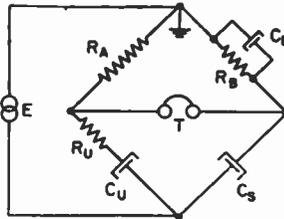


FIG. 59. Schering bridge.

Any bridge, in which the resistive balance is made by adding capacitance across a ratio arm, is now called a *Schering bridge* regardless of the position of the ground or the generator connections. If the junction of the capacitance arms is grounded, it is called an *inverted Schering bridge*. When the generator is connected across the ratio arms, it is called a *conjugate Schering bridge*.

Since the Schering bridge measures series capacitance and resistance, as indicated by Eq. (32), independent of frequency, it can be used at radio frequencies for measuring series resistance and reactance by a series substitution method.¹ The circuit of Fig. 59 is first balanced by means of C_B and C_U . The unknown impedance is then inserted in series in the U arm and the bridge rebalanced by the same controls. The scale of C_B is calibrated in ohms resistance and that of C_U in ohms reactance at a chosen frequency. The changes in readings of these scales are the series resistance and reactance of the unknown. The upper frequency limit is set by the series inductance of C_B . It is about 60 Mc for a resistance of 1,000 ohms and 140 Mc for 200 ohms.

64. Direct Capacitance. Any capacitor having terminal capacitances to a surrounding shield or to ground may be represented as a three-terminal capacitor, as

¹ Sinclair, D. B., *Proc. IRE*, 28 (11), 497-503.

shown in Fig. 60. The capacitance C_D between the terminals 1 and 2 is called the *direct capacitance*. The total capacitance between these terminals is the sum of the direct capacitance C_D , and the two terminal capacitances C_{T1} and C_{T2} in series. C_D

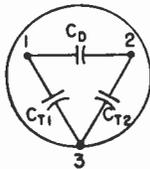


FIG. 60. Three-terminal capacitor.

may be measured on a bridge by connecting the shield to either of the junction points of the bridge, to which the direct capacitance is not connected. These two connections are shown in Fig. 61. Errors due to placing the terminal capacitances across the bridge arms greatly limit the usefulness of these connections. When the shield is connected to the junction of the ratio arms, the terminal capacitance C_{T1} is placed across the arm A and produces an error $R_A \omega C_{T1}$ in the determination of the dissipation factor of the direct capacitance C_D . The terminal capacitance C_{T2} and any capacitance of the shield to ground are placed across the detector T . When the shield is connected to the junction of the arms B and S , C_{T2} is placed across the impedance arm S and produces an error in the determination of C_D unless the standard capacitance C_S is very large compared to C_{T2} . Any capacitance of the shield to ground is also placed across C_S , while C_{T1} is placed across the generator E . If C_D is not surrounded by a shield, C_{T1} and C_{T2} are to ground, and neither of these methods is applicable.

65. Guard Circuit. The use of a guard circuit enables both direct capacitance and its dissipation factor to be measured correctly, because the terminal capacitances are not connected across any of the bridge arms. A Schering bridge with guard circuit and shielded output transformer is shown in Fig. 62. Terminal capacitance C_{T1} is placed across the guard capacitance C_H , while the other terminal capacitance C_{T2} couples the guard circuit to the junction of the bridge arms A and C_D . The standard capacitor C_S is also a three-terminal capacitor. The advantages of this construction are that all losses in the insulating supports can be carried to the guard circuit and that no capacitance will be added across ratio arm B . The guard capacitance of this capacitor is thus placed across guard capacitance C_H . Frequently this capacitance and the capacitance C_{T1} make up C_H entirely, and it becomes unnecessary to provide an extra high-voltage capacitor.

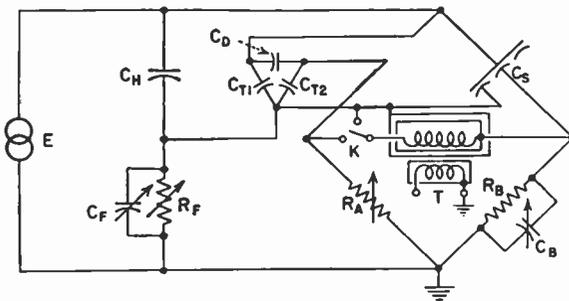


FIG. 62. Schering bridge with guard circuit.

arise when the coupling circuit is completely balanced, as is indicated by Eq. (35).

The transformer has the third shield mentioned in Sec. 61; consequently no ground capacitances are placed across the ratio arms. Instead the capacitance between the third shield and the bridge winding shield couples the guard circuit to the junction of the bridge arms B and S .

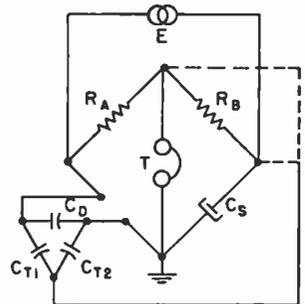


FIG. 61. Measurement of direct capacitance.

When the dielectric losses in these terminal capacitances placed across the guard circuit are excessive, it may be impossible to bring the guard circuit and bridge into balance because C_A and C_B are not large enough. The equivalent of a negative capacitance in parallel with C_F can be introduced by connecting an inductance in series with R_F , as suggested in Sec. 63. This difficulty does not

Because of the existence of the capacitances coupling the guard circuit to the bridge, the conditions of balance of the bridge involve the balancing of the guard circuit.

$$\frac{Z_A}{Z_U} = \frac{Z_B}{Z_S} = \frac{Z_F}{Z_H} \tag{34}$$

This is done by disconnecting one terminal of the output transformer from the bridge by means of switch *K* and transferring it to the guard circuit. The new bridge circuit formed by the arms *B, S, F, H* is then balanced to satisfy the right half of Eq. (34), by adjusting the guard circuit. Successive balances of bridge and guard circuits must be made until both parts of Eq. (34) are satisfied. The accuracy with which the guard circuit must be balanced in order that no appreciable error is introduced into the bridge balance depends both upon the magnitude of the coupling capacitances between the guard circuit and the bridge and also upon the degree with which they bear the same ratio to each other as *C_S* and *C_D*. The circuit formed by these coupling capacitances is called the *coupling circuit*. Its relation to the guard circuit is shown in Fig. 63. By definition, the guard circuit is that circuit which is connected across the generator, while the coupling circuit is connected across the detector. Either circuit can, therefore, be composed of similar or dissimilar elements. The circuit devised by Wagner in 1911 for the same purpose and called a *Wagner ground* was always composed of similar elements and connected across the ratio arms, where the generator was also connected. By the above definition it was a guard circuit.

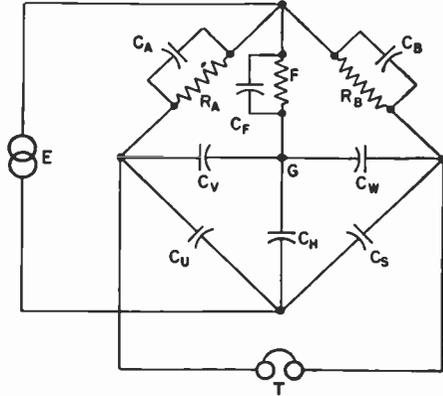


FIG. 63. Schering bridge with guard circuit and coupling circuit.

Balsbaugh¹ has shown that for the network of Fig. 63 the conditions of balance are either those given in Eq. (34) or those given in Eq. (35).

$$\frac{Z_A}{Z_B} = \frac{Z_U}{Z_S} = \frac{Z_V}{Z_W} \tag{35}$$

If either the guard circuit or the coupling circuit is partially balanced, the other circuit need be only partially balanced in order to introduce no appreciable error in the bridge balance. Balance of both circuits may be conveniently made without disconnecting either generator or detector by connecting their junction *G* to those corners of the bridge which place the guard or coupling circuit in parallel with similar elements of comparable impedance. Ground may be placed at any corner of the bridge or at the junction *G* of guard and coupling circuits. This latter point is best in many respects because it simplifies the mechanical construction of the bridge and avoids the need of any insulated shields. All elements used in the bridge must, however, be three-terminal impedances. Grounding the point *G* also simplifies the construction of the three-terminal measuring cell by making it unnecessary to provide an insulated shield inside the outer grounded case.

66. Comparison of Inductances and Capacitances. An inductance and a capacitance may be compared directly by suitably placing them in the four-impedance network. The connections for the Maxwell bridge are shown in Fig. 64. The balance

¹ Balsbaugh and Herszenberg, *J. Franklin Inst.*, 218 (1), 49-97.

equations are

$$L_U = R_A R_S C_B \quad \text{and} \quad R_U = \frac{R_A}{R_B} R_S \quad (36)$$

whence

$$Q_U = Q_B$$

Losses in C_B enter only into the resistance balance and may be made negligible by suitable choice of R_A . The resistance and reactance balances are not independent unless C_B is continuously variable or resistance is added in series with the unknown inductor.

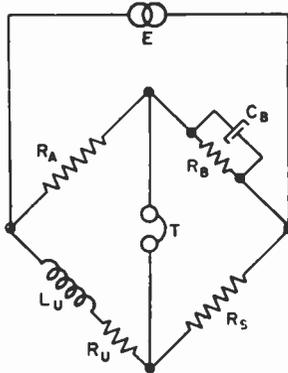


FIG. 64. Maxwell bridge.

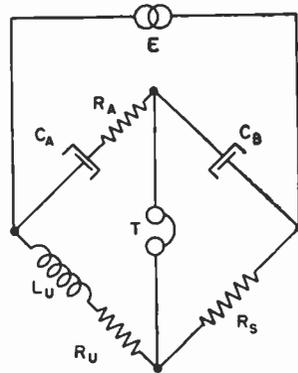


FIG. 65. Series Owen bridge.

In the series Owen bridge an inductance is compared with a capacitance in the manner shown in Fig. 65. The balance equations are

$$L_U = R_A R_S C_B \quad \text{and} \quad R_U = \frac{C_B}{C_A} R_S \quad (37)$$

whence

$$Q_U = Q_A$$

The resistance balance is made either by having C_A continuously variable or by adding resistance in series with the unknown inductor.

Both of these bridges measure series inductance and resistance. The parallel inductance and resistance may be calculated from

$$L_P = L_S(1 + D^2) \quad \text{and} \quad R_P = R_S(1 + Q^2) \quad (38)$$

They are both subject to errors caused by shunt capacitances across the resistive arms, such as were outlined in Sec. 62. The effect of these shunt capacitances is negligible for inductance except when the storage factor Q_U approaches unity. They are large for resistance, particularly when the storage factor Q_U is large, and set an upper frequency limit of a few kilocycles in many cases.

In the parallel Owen bridge an inductance is compared with a capacitance in the manner shown in Fig. 66. The balance equations are identical with those given for the series Owen bridge in Eq. (37).

The Hay bridge, shown in Fig. 67, may be considered to be a parallel Maxwell bridge. The balance equations are identical with those given for the Maxwell bridge in Eq. (36).

Both of these bridges measure parallel inductance and resistance. The series inductance and resistance may be calculated from the expressions

$$L_s = \frac{L_P}{1 + D^2} \quad \text{and} \quad R_s = \frac{R_P}{1 + Q^2} \quad (39)$$

They are both subject to the same errors caused by shunt capacitances across the resistive arms as have been outlined previously for the series bridges.

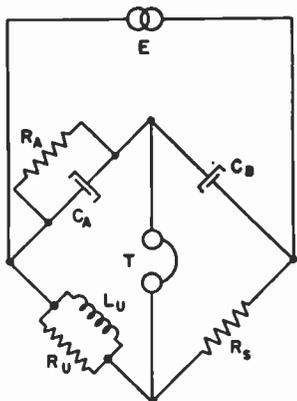


FIG. 66. Parallel Owen bridge.

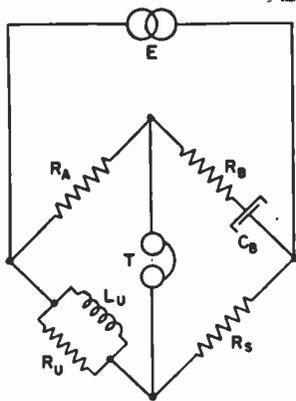


FIG. 67. Hay bridge.

When the incremental inductance of an iron-cored inductor is measured, it is necessary to send a d-c magnetizing current through the inductor. The series Owen or Hay bridge is usually chosen because the capacitance C_B in the B arm blocks the d-c magnetizing current from one side of the bridge and allows this current to be measured externally.

67. The resonance bridge shown in Fig. 68 is the simplest bridge in which inductance, capacitance, and frequency enter. At balance, the arm containing the reactances is resonated to the applied frequency and becomes a pure resistance. The bridge can then be made an all-resistance equal-arm bridge. For this reason it may be used at high frequencies to measure the resistance and inductance of an inductor.

The balance equations are

$$\omega^2 = \frac{1}{L_U C_U} \quad \text{and} \quad R_U = \frac{R_A}{R_B} R_S \quad (40)$$

The resistance measured is the series resistance of the inductor.

It is equally possible to place the unknown inductor and capacitor in parallel. Equation (40) still holds except that R_U will be the parallel resistance of the inductor. This bridge is frequently used to measure frequency, usually in the a-f range. A variable inductor is used, and the capacitor may be varied in steps. A range from 200 cycles to 4 kc may be covered in three ranges. The frequency scale is irregular, owing to the characteristics of variable inductors, and the various ranges cannot be made multiples of one another. Owing to the large stray field of the variable inductor, its magnetic pickup is considerable. A resistance balance must be provided to allow for the variation of the resistance of the resonance arm with frequency.

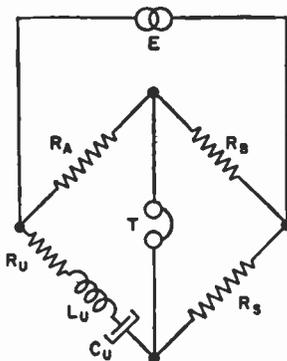


FIG. 68. Resonance bridge.

68. Wien Bridge. Capacitances may be measured in terms of resistance and frequency with the Wien bridge, shown in Fig. 69. The balance equations expressed in their simplest form are

$$\omega^2 = \frac{1}{R_U R_S C_U C_S} \quad \text{and} \quad \frac{C_U}{C_S} = \frac{R_B}{R_A} - \frac{R_S}{R_U} \quad (41)$$

Solving for the two capacitances,

$$C_U^2 = \frac{R_B R_U - R_A R_S}{R_A R_S R_U^2 \omega^2} \quad \text{and} \quad C_S^2 = \frac{R_A}{(R_B R_U - R_A R_S) R_S \omega^2} \quad (42)$$

The bridge is valuable because the standards of frequency and resistance are known to a greater accuracy than the standard of capacitance. Ferguson and Bartlett¹ have developed this method to its greatest precision. Their estimated accuracy for the determination of capacitance by this method is 0.003 per cent.

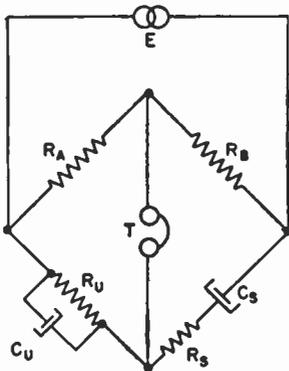


FIG. 69. Wien bridge.

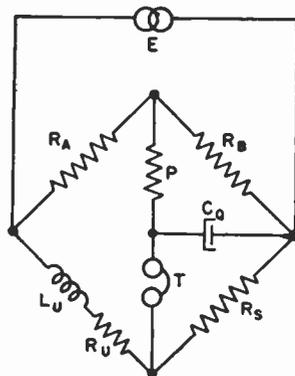


FIG. 70. Anderson bridge.

The Wien bridge also furnishes a very convenient means for measuring frequency in the a-f range. The two capacitances are made equal, while the two ratio arms are made such that R_B is twice R_A . The two resistances R_U and R_S are made variable over a suitable range but are also kept equal. Thus the resistance balance is always satisfied and the reactance balance reduces to

$$f = \frac{1}{2\pi R_U C_U} \quad (43)$$

In a commercial frequency meter R_U and R_S are wound on tapered cards so shaped that the frequency scale is logarithmic. This gives a constant fractional accuracy of reading. There are three frequency ranges, obtained from three different pairs of capacitors, each covering a range of 10 to 1 in frequency. The same calibration serves for all ranges. The frequency limits attained are 20 cycles and 20 kc.

69. Six-impedance Network. The six-impedance network was developed by Anderson to provide a modification of the Maxwell bridge which would render the two balance conditions independent even with a fixed capacitance. The connections are shown in Fig. 70.

The general balance condition for the six-impedance network is

$$Z_Q(Z_B Z_U - Z_A Z_S) = Z_P[Z_P(Z_A + Z_B) + Z_A Z_B] \quad (44)$$

¹ Ferguson and Bartlett, *Bell System Tech. J.*, 7 (3), 420-437.

For the Anderson bridge this reduces to

$$L_U = \left[R_S C_Q R_P \left(1 + \frac{R_A}{R_B} \right) + R_A \right] \quad \text{and} \quad R_U = \frac{R_A}{R_B} R_S \quad (45)$$

The effect of losses in C_Q is usually small.

70. Mutual-inductance Balances. Two mutual inductances may be compared by means of the *Felici mutual-inductance balance* shown in Fig. 71. The known mutual inductance must be variable. For the usual condition of balance, zero voltage across the null detector, the two mutual inductances are equal.

$$M_U = M_S \quad (46)$$

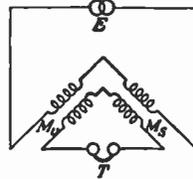


Fig. 71. Felici mutual-inductance balance.

They must be so connected that their induced secondary voltages are in opposition. Mutual inductance between them should be avoided.

71. Four-impedance Network with Mutual Inductances.

A mutual inductance may be compared with a self-inductance on a four-impedance bridge by placing it between one arm and either an input or output lead of the bridge, as shown in Fig. 72.

The general balance equation for this network is

$$Z_A Z_S - Z_B Z_U - j\omega M(Z_A + Z_B) = 0 \quad (47)$$

For Campbell's arrangement of this bridge the two conditions of balance become

$$L_U = \frac{R_A}{R_B} L_S - \left(1 + \frac{R_A}{R_B} M \right) \quad \text{and} \quad R_U = \frac{R_A}{R_B} R_S \quad (48)$$

A substitution method is usually adopted so that the inductance and resistance of that portion of the mutual inductance in the S arm need not be known. When the ratio arms are equal, the extra balancing inductance represented by L_U of Fig. 72 may be

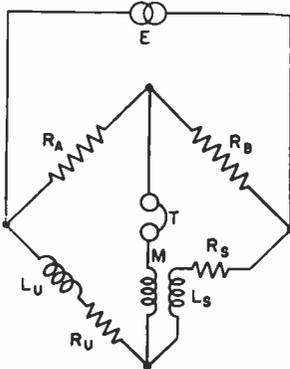


Fig. 72. Comparison of mutual with self-inductance.

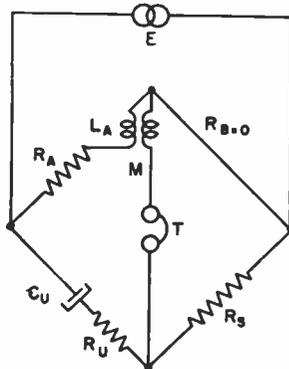


Fig. 73. Carey Foster mutual-inductance bridge.

eliminated by providing a center tap in one branch of the mutual inductance. This connection is usually referred to as the Heaviside equal-arm bridge.

A mutual inductance may be compared with a capacitance by means of the *Carey Foster bridge*, shown in Fig. 73. The conditions of balance are

$$C_U = \frac{M}{R_A R_S} \quad \text{and} \quad R_U = R_S \left(\frac{L_A}{M} - 1 \right) \quad (49)$$

The impedance of the *B* arm is made zero in order to make the balance independent of frequency. The method suffers because the resistance and self-inductance of the mutual inductance enter into the expressions for the unknown capacitance and its resistance, respectively. Capacitance between the two windings of the mutual inductance causes the voltage induced in its secondary to have a phase angle with reference to the primary current different from 90 deg. This reduces the calculated resistance of the capacitor and frequently yields negative values, especially for large mica capacitors. The method is perhaps better suited for the measurement of a mutual inductance in terms of a known capacitor.

T NETWORKS

72. Two or more T networks connected in parallel provide a method of null balance which in many respects is equivalent to an a-c bridge circuit.¹ The connections for two T networks are shown in Fig. 74. The most important feature of the network is that generator and detector have a common terminal, which can be grounded. Hence no shielded transformer is necessary. This is a considerable convenience at low frequencies and also makes it possible to use the network at high frequencies up to at least 30 Mc.

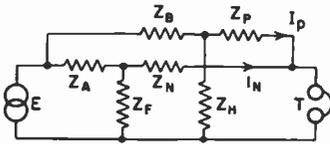
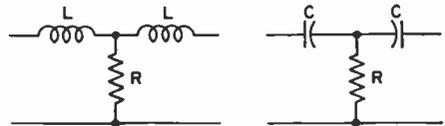


Fig. 74. Parallel T network.



$$R_T = -\frac{\omega^2 L^2}{R}$$

$$X_T = 2\omega L$$

$$R_T = -\frac{1}{R\omega^2 C^2}$$

$$X_T = \frac{2}{\omega C}$$

Fig. 75. T networks having negative transfer resistance.

The condition for a null deflection of the detector is that the currents in the output circuits of the two networks shall be equal and opposite.

$$I_P + I_N = 0 \tag{50}$$

These currents are best evaluated by considering the transfer impedances, which are defined as the ratios of the input voltage to the output current.

$$Z_{TP} = \frac{E}{I_P} = Z_A + Z_N + \frac{Z_A Z_N}{Z_F}$$

$$Z_{TN} = \frac{E}{I_N} = Z_B + Z_P + \frac{Z_B Z_P}{Z_H} \tag{51}$$

Hence

$$Z_A + Z_N + \frac{Z_A Z_N}{Z_F} + Z_B + Z_P + \frac{Z_B Z_P}{Z_H} = 0 \tag{52}$$

Under somewhat restricted conditions this equation can be satisfied because the impedances are complex quantities. Although any of the terms of Eq. (52) may contain negative reactances, only the product terms can have a negative resistance. The only two T networks having a negative resistance component of transfer impedance are shown in Fig. 75. One of these networks or a modification must be used in every parallel T network which can be balanced.

73. Parallel T Networks. The parallel T network shown in Fig. 76 is equivalent to the Wien bridge² and has similar balance equations.

¹ Tuttle, W. N., *Proc. IRE*, 28 (1), 23-29.

² See Sec. 68.

$$\omega^2 = \frac{C_B + C_P}{R_A R_N C_B C_P C_F} \quad \text{and} \quad \frac{C_B + C_P}{C_F} = \frac{1}{R_H} \frac{R_A R_N}{R_A + R_N} \quad (53)$$

When both the T networks are made symmetrical and when in addition C_F is made twice C_B and R_A is made twice R_H , the resistance balance is always satisfied and the reactance balance reduces to

$$f = \frac{1}{2\pi R_A C_B} \quad (54)$$

which is identical in form with Eq. (43).

7A. Bridged T Networks. When the shunt arm of one of the T networks is made infinite, the circuit is called a *bridged T network*. The circuit shown in Fig. 77 is very convenient for measuring inductance in terms of capacitance, resistance,

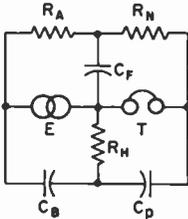


FIG. 76. Parallel T network equivalent to Wien bridge.

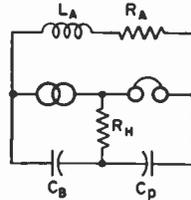


FIG. 77. Bridged T network.

and frequency. It is equivalent to the resonance bridge (Sec. 67). The balance equations are

$$L_A = \frac{2}{\omega^2 C_B} \quad \text{and} \quad R_A = \frac{1}{R_H \omega^2 C_B^2} \quad (55)$$

whence

$$Q_A = 2R_H \omega C_B$$

At balance the full generator voltage appears across the inductor. When the junction of generator and detector is grounded, the terminal capacitances of the inductor are placed across generator and detector, and the direct impedance of the inductor is measured. For iron-cored inductors the d-c magnetizing current can be conveniently introduced in series with the a-c voltage, provided a low resistance path is placed across the detector.

References

Bolton, E. P.: "Electrical Measuring Instruments," Dutton, 1923.
 Brown, H. A.: "Radio-frequency Electrical Measurements," McGraw-Hill, 1938.
 Cambridge Instrument Company, Catalogue.
 General Radio Company, Catalogue.
 Hague, B.: "Alternating-current Bridge Methods," Pitman, London, 1938.
 Hund, A.: "High-frequency Measurements," McGraw-Hill, 1933.
 Jansky, C. M.: "Electrical Meters," McGraw-Hill, 1913.
 Kennelly, A. E.: "Electrical Vibration Instruments," Macmillan, 1923.
 Laws, F. A.: "Electrical Measurements," McGraw-Hill, 1938.
 Leeds & Northrup Company, Catalogue.
 Moullin, E. B.: "Radio Frequency Measurements," Griffin, 1926.
 Starling, S. G.: "Electricity and Magnetism," Longmans, 1924.
 Terman, F. E.: "Measurements in Radio Engineering," McGraw-Hill, 1935.
 Weston Electrical Instrument Corporation, Catalogue.

CHAPTER 7

POWER-SUPPLY SYSTEMS

BY RICHARD C. HITCHCOCK¹

1. Direct-current Power Requirement. The electrical power required for operating radio transmitters and receivers is usually "steady" d.c. for plate and grid circuits. Depending on conditions, either d.c. or a.c. is employed for heating tube filaments or cathodes. This chapter deals mainly with the rectifiers and filters for supplying the desired d.c. In succeeding paragraphs typical radio sets and amplifiers are described briefly, after which the component parts and their relations are analyzed.

2. Type of Service. The type of supply depends largely on the amount of power to be furnished. A portable receiver may operate for some time from self-contained dry-cell batteries, but a 50-kw broadcast transmitter on a regular schedule requires considerably more energy. An explorer can operate a hand-cranked generator of 50 or even 100 watts capacity for a short time, but for longer periods other devices are more applicable.

Power supplies can be divided into two types: one for regular use and the other the emergency set or stand-by. The latter is ready on a few seconds notice and is capable of supplying sufficient energy for regular operation. Preferably two independent sources of a-c power are provided for multikilowatt stations, on either of which the station can radiate full power. A gasoline-electric set may serve as one of these sources, being independent of long line wires.

CHARACTERISTICS AND MEASUREMENT OF D.C.

3. Indicating Instruments. Since d.c. is employed largely for radio transmitters and receivers, a brief analysis will be made of its various kinds and their measurement. One reason for this analysis is that instruments of the repulsion-iron or dynamometer type will not read the same as an "average" type on certain kinds of d.c. This difference sometimes causes confusion.

If a d-c ammeter is specified, it usually refers to a D'Arsonval instrument (permanent magnet, moving coil), one that reads "average" values.

Figure 1 shows five typical kinds of d.c., one or more of which are present in any d-c power supply. Steady d.c. is the output from a primary or secondary (storage) battery, or from a suitable filter connected to a pulsating or ripple d-c source. Ripple d.c. is the usual output from a d-c generator, the ripple being a result of commutation. Three types of pulsating d.c. are (1) half-wave rectified single-phase (see also Fig. 24); (2) full-wave rectified single-phase (Fig. 37); and (3) three-phase full-wave rectified (Fig. 13).

The ambiguity of the term "direct current" is readily apparent when considering Fig. 1, since all these wave forms fall into this classification; *i.e.*, each remains positive. None goes to a negative value. The data on the figure show the minimum voltage as a decimal part of the maximum voltage E_m ; *e.g.*, 65 per cent of its maximum, for ripple d.c.

The second column shows the average value of the potential difference as a factor times the maximum E_m . The factor for pulsating d.c. varies from 0.32 for the half-wave rectified single-phase to 0.96 in the case of the full-wave six-phase.

The rms or effective value of a current is such that the heating effect (I^2R) is the

¹ Westinghouse Research Laboratories.

same for d.c. or a.c., by definition. For pulsating d.c. the wattage found by multiplying the average voltage by the average current is not the same as the rms voltage times the rms current. The results of these average readings are sometimes called¹ "d-c watts."

The readings of different types of instruments can be predicted from the value of the form factor. On the ripple d.c. from a d-c generator, when the form factor is 1.01 as shown, a D'Arsonval instrument would read 1 per cent lower than would a dynamometer type of instrument. The D'Arsonval instrument reads average, and the dynamometer reads rms. When there is a difference in readings, the rms instrument always reads higher. For the pulsating d-c output of the single-phase half-wave rectifier, the form factor is 1.57, and an rms instrument (repulsion-iron, dynamometer, or thermo-couple) would read 57 per cent higher than the average-reading instrument (D'Arsonval).

Power transformers for radio receivers and amplifiers are usually rated by specifying d-c volts and d-c amperes output. However, to design a transformer which has rms a.c. flowing in its windings to supply a filter which furnishes average d.c. to the set, the form factor is employed. A full-wave rectifier has the characteristic form factor 1.11, which means that this is the ratio of the rms to average. To supply 100 ma average, the transformer furnishes 111 ma rms (55 ma from each secondary). See Sec. 29 for a specific design.

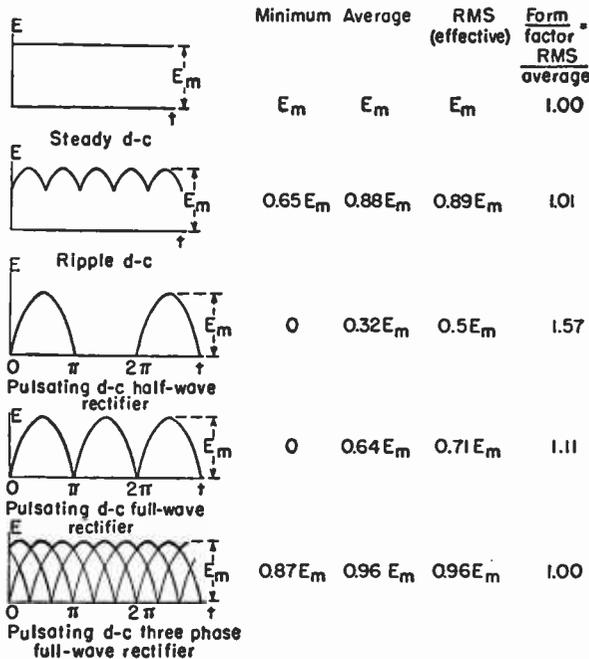


FIG. 1. Types and characteristics of d.c.

TYPICAL POWER-SUPPLY SYSTEMS

4. Television Receiver Power Supply.² A 10,000-volt d-c power supply for a direct-vision television receiving tube is shown in Fig. 2. One half of the 6SN7 tube is a blocking oscillator, the 807 is a pulse tube, and the 8016 is a rectifier. The blocking oscillator repetition rate (due to R_1, C_3) is much higher than the horizontal line frequency of 15.75 kc. A safety feature is that the cathode is biased to cutoff, no oscillation being possible unless a triggering signal is supplied through C_2 . This triggering signal is a positive saw tooth, at 15.75 kc.

The pulse from the 6SN7 on the 807 grid cuts off plate current at the 15.75-kc rate, the plate current gradually increasing to a finite value between pulses.

Inductance T_2 lowers the peak voltage on the 807, while maintaining a high available voltage. The exact position of the tap on T_2 is important. This is approxi-

¹ Smith, I. R., Rectox Rectifier Testing, *Elec. J.*, August, 1938, p. 328.

² Allen B. Du Mont Laboratories, Inc., Passaic, N.J.

117N7GT filament is heated across the line, and the rectifier section of this tube supplies half-wave energy to the filter choke. The other filaments in series are heated by the d-c plate current of the output pentode section of the 117N7GT, which also provides bias potential, and this connection furnishes bias voltage for the control grid of the power tube. Here the battery output tube 3Q4 is not used, and its filament is not heated since the Y connection in the female receptacle is not grounded.

When battery power is used, the plug XY is inserted in the female receptacle in the set, the Y connection bringing into the circuit the filament of the 3Q4 by grounding the negative terminals of the 9-volt A battery. This Y connection also connects the negative terminal of the 90-volt B battery to ground. The filament of the 117N7GT is not lighted on battery power.

6. Battery-operated Receiver. Figure 5 is a completely battery-operated radio receiver¹ which uses a 1.5-volt A battery for the five tube filaments (parallel), and two series-connected 45-volt B blocks. Characteristic of both Figs. 4 and 5 is the permanent magnet moving-coil loud-speaker which does not require external power for energizing its magnetic circuit.

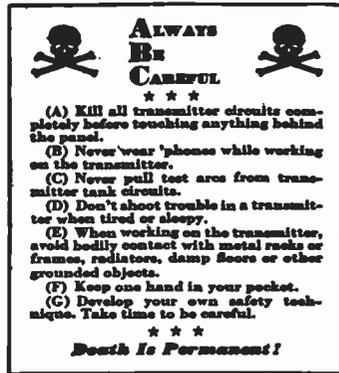


FIG. 3. A good sign to be posted in high-voltage laboratory.

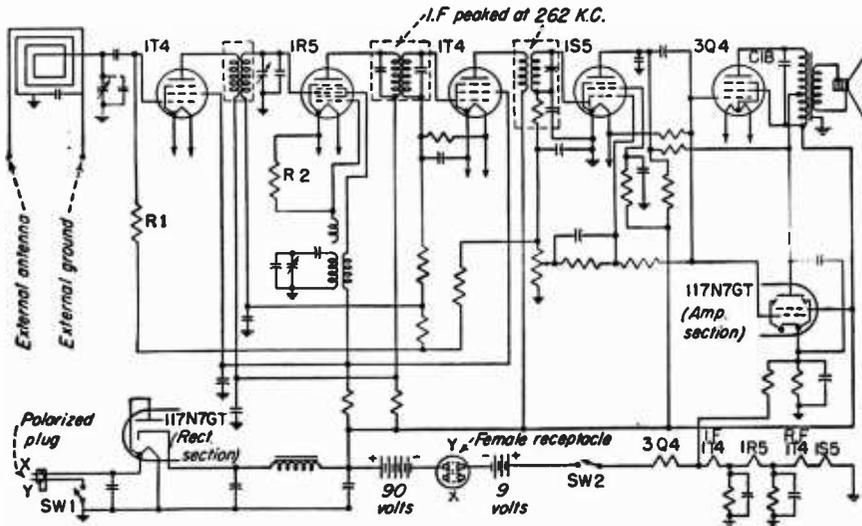


FIG. 4. Universal (a.c., d.c., battery) receiver circuit (Emerson Model 505).

7. Receiver for A.C.-D.C. Figure 6 is typical of a portable a-c-d-c receiver.² Each tube heater takes 0.15 amp from the line, and there is a tap on the rectifier tube filament 35Z5GT to supply the pilot light. The output tube is 50L6GT. The tubes light up whenever turned on, but, as with all a-c-d-c models, used on d.c., it is neces-

¹ Emerson Radio & Phonograph Co., Model CX.
² Emerson Radio & Phonograph Co., Model 507.

sary to insert the power plug correctly to get operation. The a-c plug may be inserted for minimum hum.

8. Receiver Power Unit for 115 to 230 Volts D.C. Figure 7 shows a vibrator-transformer-rectifier circuit¹ which has two ranges of d-c power input, 105 to 125, and

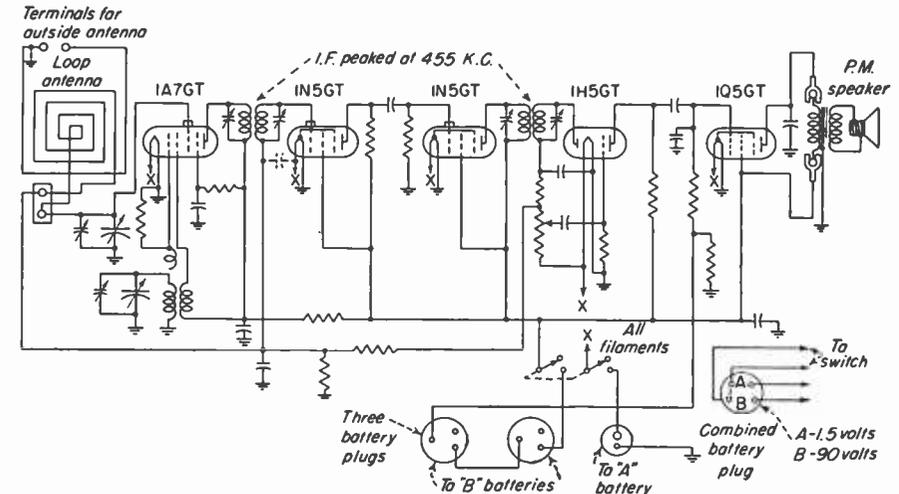


FIG. 5. Battery-operated receiver circuit.

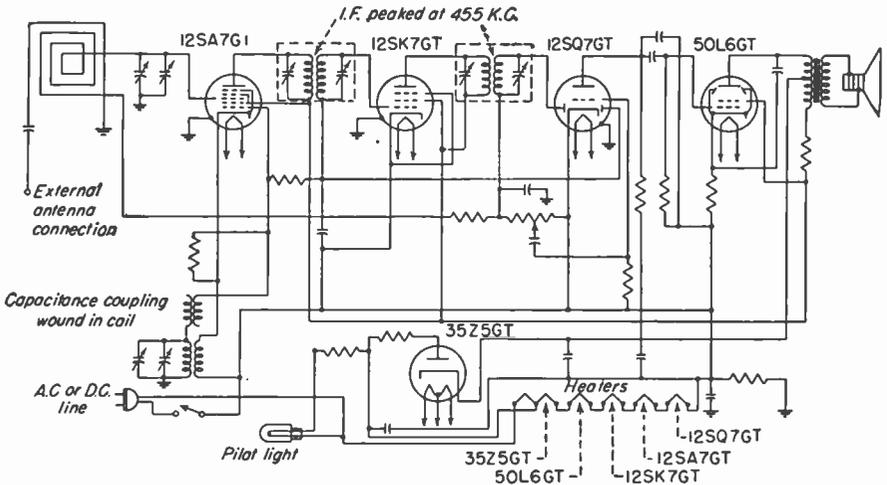


FIG. 6. Circuit for typical a-c-d-c receiver.

210 to 250 volts. The link board shown in the figure changes the unit to suit the power voltage available. These units provide an a-c heater voltage of 6.3 volts and rectified plate potential of 360 to 400 volts.

9. Receiver Power Units for 6 Volts D.C. and Various A.C. Figure 8 shows the two different units required when a radio receiver is to be operated either on 6 volts

¹ RCA Model PSU 10E, p. 312, *Service Notes*, 1938.

phase transformer, and the plate circuit comprises a three-phase transformer with a double-Y secondary, fed by a delta-connected three-phase primary.

Means are generally provided for placing a spare rectifier in the circuit. The spare tube filament is kept lighted, and its plate lead wired to all six inactive jaws of a rack of 6 spdt switches. The blade of each switch is a transformer lead, and the second jaw of each switch goes to the rectifier tube in use. If a rectifier fails, the switch is thrown (either automatically or by hand) from the regular tube plate to that of the spare tube, thus making the spare tube active and taking plate voltage off the regular

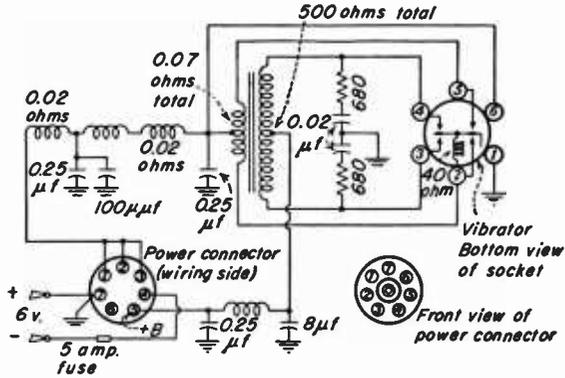


FIG. 8a. Power supply from 6 volts d.c.

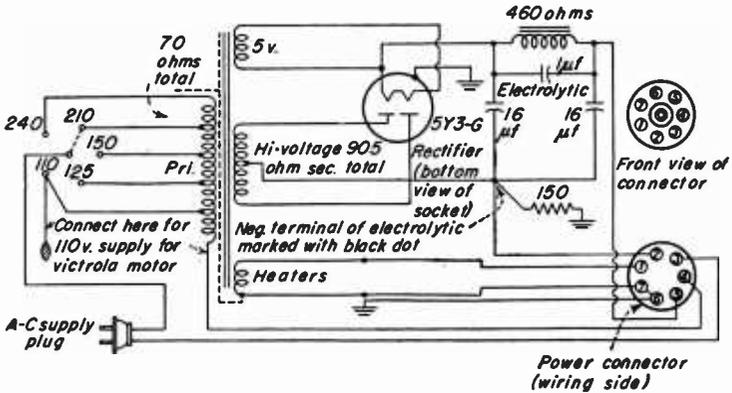


FIG. 8b. Power supply from a.c. of 110 to 240 volts.

tube, so that it can be replaced and then reenergized when the switch is thrown back to its original position.

By using a suitably connected three-phase filament transformer, some increase in the life of hot-cathode mercury rectifier tubes can be obtained. The plate circuits are the same as shown in Fig. 12, but the filament supply of any one tube is connected so that it is 90 deg out of phase with the plate voltage supplied to it. This makes the filament a unipotential device the moment the crest plate current is drawn.

13. Vacuum-type Rectifiers for Receivers. Typical rectifier tube characteristics for receiving sets and medium power audio amplifiers are shown in Figs. 14 to 17. For each tube the d-c output volts available at the filter input are shown as a function

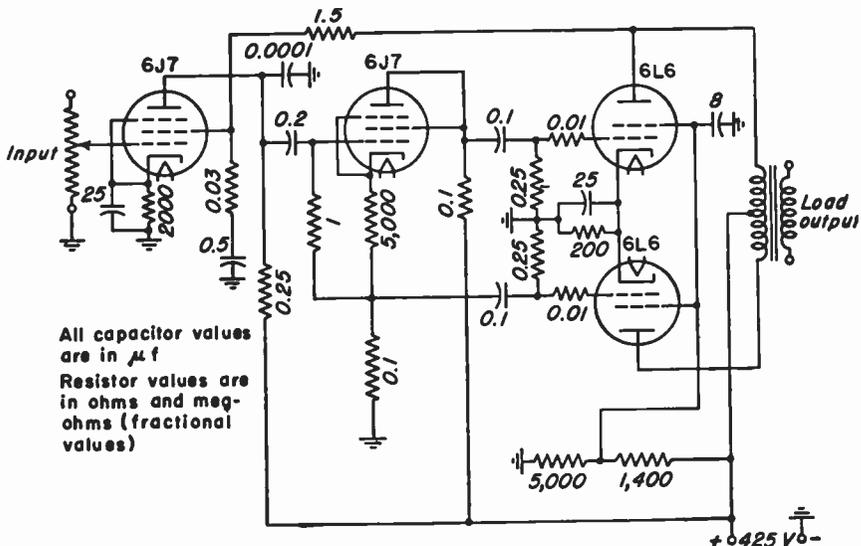


FIG. 9. Audio amplifier with 32-watt output.

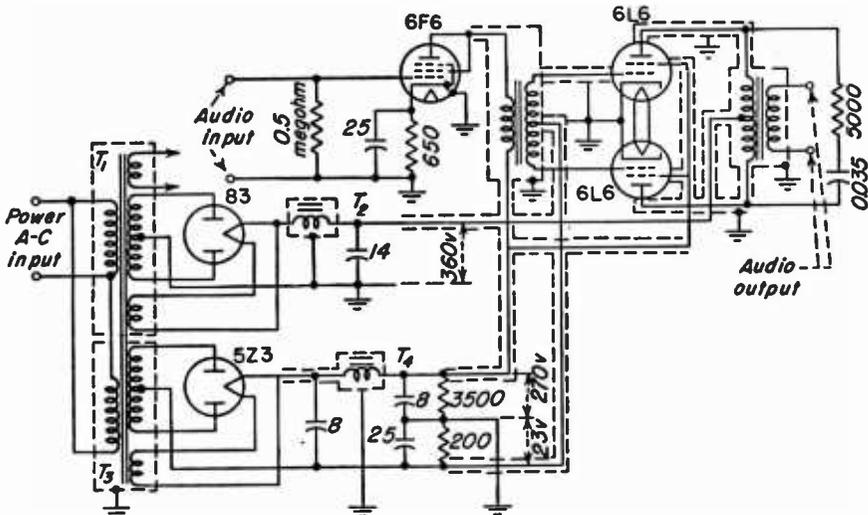


FIG. 10. Power supply, driver and output stage of 45-watt-output, class AB₂ audio amplifier.

of output current drain. A detailed analysis of the application of such tubes, as well as an excellent graphical analysis, can be found in the reference.¹

Taking the tube type 1-v as typical, Fig. 14 gives the relation between the d-c output volts at the input to the filter and the current taken from the system. The curves show that a higher voltage is secured by use of a larger input capacitor (see Sec. 30 for filters). Thus the curve shows that with an input of 117 volts a.c. a load of 30 ma

¹ Schade, Otto, Analysis of Rectifier Operation, *Proc. IRE*, 31 (7), 341-361, July, 1943.

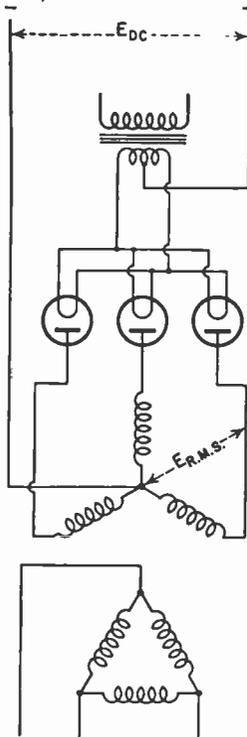


FIG. 11. Three-phase, half-wave.

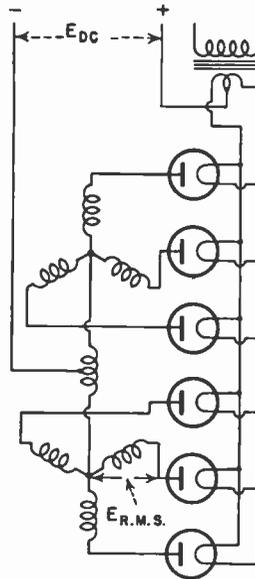


FIG. 12. Three-phase, double-Y.

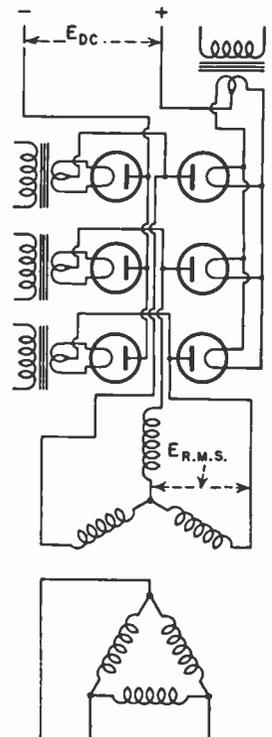


FIG. 13. Three-phase, full-wave.

Polyphase Rectifier Filter Circuits

Rectifier circuit	Three-phase half-wave	Three-phase double-Y	Three-phase full-wave
E_{dc}	$1.170 E_{rms}$	$1.170 E_{rms}$	$2.34 E_{rms}$
E_{max} (inverse)	$0.827 E_{max}$	$0.827 E_{max}$	$1.65 E_{max}$
I_{avg}	$2.09 E_{avg}$	$2.09 E_{avg}$	$1.045 E_{avg}$
Ripple frequency	$0.827 E_{max}$	$1.91 I_{max}$	$0.955 I_{max}$
Ripple voltage (rms)	$3f$	$6f$	$6f$
Secondary va	18.3%	4.2%	4.2%
Primary va	$1.48 E_{dc}/I_{dc}$	$1.48 E_{dc}/I_{dc}$	$1.05 E_{dc}/I_{dc}$
	$1.21 E_{dc}/I_{dc}$	$1.05 E_{dc}/I_{dc}$	$1.05 E_{dc}/I_{dc}$

d.c. can be drawn at an output of 123 volts d.c. Figure 15 gives the average plate characteristic, i.e., the voltage drop through the tube. At 30-ma drain, the voltage drop in the tube is about 9 volts.

There are several equivalent tubes with different numbers and bases as shown in Table 1.

Table 1. Rectifier Tubes with Similar Characteristics

5T4, metal octal	5Y3GT, glass octal	25Z5, 6-pin socket
5U4G, glass octal	5Y4G, different base	25Z6, glass octal
5X4G, different base	5Z4, metal octal	50Y6GT, 50-volt heater
5Z3, glass, 4-pin socket	80, glass, 4-pin socket	

14. Voltage-doubling Circuits.¹ The typical full-wave voltage doubler of Fig. 21 has output characteristics of Fig. 18 for a 25Z6 tube. Note that neither side of the d-c output can be grounded, since the input a.c. (one side grounded) is at a different potential.

The half-wave doubler of Fig. 20 has one d-c lead at the same potential as the line, and this can be used for high-gain receivers and amplifiers (McIlvaine patent).

Figure 22 shows the comparative regulation of the three circuits, Figs. 19 to 21, when using the same total capacitance, 12 μ f, and the same input a-c voltage, 117.

Figures 19 and 20 have a common d-c and a-c terminal, but the half-wave rectifier has lower output voltage, with the compensating advantage of better output voltage regulation.

Electrolytic capacitors are normally used for the circuits of Figs. 19 to 21, with the usual precaution of proper polarity, and an added precaution in the case of C_1 for Fig. 20: if heavy d-c loads are drawn from the circuit of Fig. 20, about 50 ma for a 25Z6, C_1 may reverse in polarity, in which case a special a-c electrolytic capacitor is required, or a paper capacitor with no special polarity.

Although Figs. 20 and 21 show separate tubes, it will be clear that a single 25Z6, 50Y6GT, or 117Z6 tube, each having two cathodes and two plates in one glass envelope, will be suitable. The precaution to observe is that of heater-cathode potential; e.g., the 25Z5 and 25Z6 have 350 volts maximum "d-c heater-cathode potential." Rectifier tubes with larger current-handling ability than 25Z5 may also be used in the

circuits of Figs. 20 and 21, even if the heater-cathode potential exceeds the allowable value, if separate tubes are used for each diode shown and separate filament heater windings are provided.

A resistor R of 50 to 100 ohms should be used, as shown in Figs. 19 to 21, when the capacitor values are greater than 8 μ f. This is required because any uncharged capacitor acts like a short circuit, and until the voltage

builds up across its terminals, a heavy charging current flows. When a cold rectifier is turned on, the charging current to a large capacitor is usually less than the allowed peak current, since the cold cathode of the tube is itself a high resistance in series with

¹ McIlvaine, O. T., U.S. Patent 1946354, 1934. Electronic Products Co., St. Charles, Ill. Hitchcock, R. C., Thermionic Rectifiers, *Electronics*, 1944, pp. 102-105, 226-230.

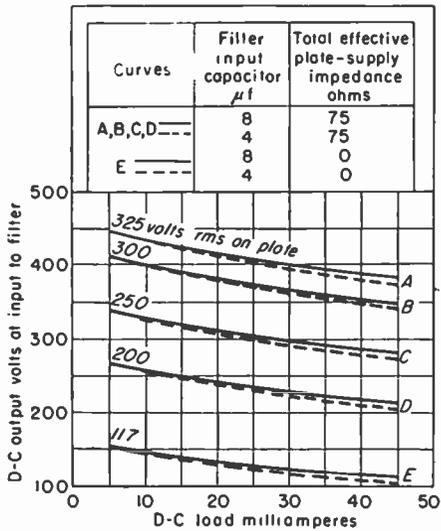


FIG. 14. Operation characteristics 1-v rectifier.

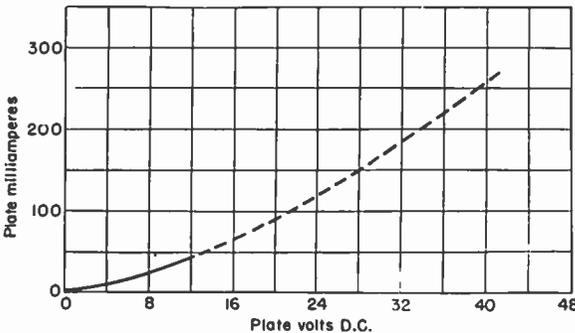


FIG. 15. Average plate characteristic, 1-v rectifier.

the capacitor. But if the switch is turned off and on in rapid succession, the still-hot cathode of the rectifier will pass a heavy current to the partly discharged capacitor. This current may be high enough in value and long enough in duration to burn off the

tube lead-in wires, if the capacitor is of the order of 40 μ f.

Dry contact rectifiers have no appreciable warm-up time. Even though their ratings are largely that of temperature (see Secs. 20 to 22) a similar series resistor is often employed to limit the surge current to some maximum value specified by the manufacturer.

Selenium stacks may be used to replace the

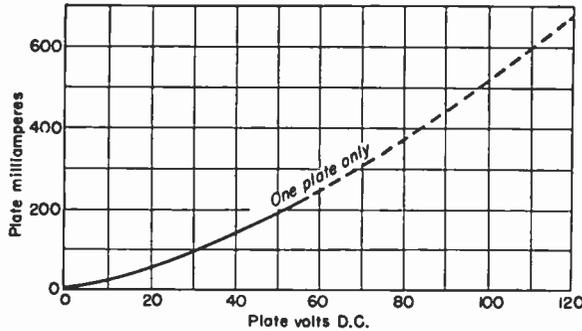


FIG. 16. Average plate characteristic, 5U4-G rectifier.

diodes of Figs. 19 to 21 (also see Fig. 24), and for these dry contact rectifiers there is no heater supply to take into account. The allowable back emf must of course be high enough to withstand the potential applied in the reverse direction (see Sec. 22).

15. Gas-filled Rectifiers for Receivers. The mercury-vapor rectifier tubes types 82 and 83 have an approximate drop of 15 volts, independent of the load. This provides good regulation, *i.e.*, small loss of potential with increasing load, but it necessitates care in using an appropriate load. A mercury-vapor rectifier will supply currents far in excess of its rating if it is connected to a load too low in impedance. Usually the load has sufficient impedance (largely resistance) to limit the current to a safe value. For example, in Fig. 10 the choke, the output transformer, and the push-pull 6L6 tubes offer an effective load to the 83 rectifier. There is a limit to the safe operating temperature of this type of tube.

The hot-cathode mercury-vapor rectifier¹ differs from the mercury-arc in two respects: (1) It operates at a relatively low temperature, so that the vapor pressure is low. This low pressure gives the useful characteristic of having a high breakdown in the inverse direction. (2) The electrons are emitted from the filament and not from a pool of mercury. In this second respect this tube resembles the vacuum-tube rectifier, but the difference lies in the much lower potential drop due to the neutralizing of the filament space charge by the positively charged mercury ions. The mercury-vapor tube is self-igniting and does not require the starting mechanism of the mercury-arc rectifier.

16. Battery Chargers. For low-voltage high-current rectification the argon-filled tungsten filament Rectigon and Tungalbulbs fill the need (see also Sec. 19). The use is largely that of charging storage batteries, and no filter is needed for this application.

¹ Pike, O. W., and H. T. Maser, *QST*, February, 1929, p. 20.

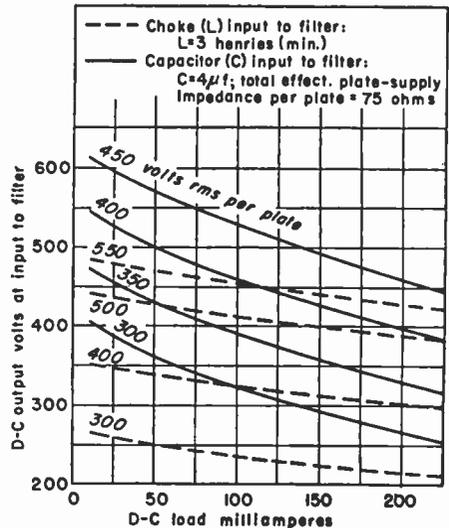


FIG. 17. Operation characteristics, 5U4-G rectifier.

Table 2. Battery-charger Tube Characteristics
(All half-wave single phase)

Number*	Filament		Max d-c anode ratings		Max inverse voltage	Over-all length, in.
	Volts	Amp	Volts	Amp		
289415	2	12	75	2	275	5½
289416	2.2	18	90	6	375	6¾
766776	2.5	27	60	15	225	8¾

* Style number of Westinghouse Electric Corp., Rectigon tubes. General Electric manufactures similar tubes under the name Tungar.

Filters have been designed for use with these rectifiers, so that the output can be fed directly to the filaments of d-c radio tubes.

To design a low-pass filter for this purpose, Eqs. (3) and (4), Sec. 30, may be used, since the chart of Fig. 33 does not cover so high a current (low resistance). The capacitor is usually electrolytic, and polarity must be observed.

17. Mercury-arc Rectifiers. Formerly mercury-arc rectifiers were most used¹ in the field lying between the argon tube and the filament vacuum rectifier. With the introduction of mercury-vapor tubes, many of the advantages of the mercury-arc rectifier (low voltage drop, high efficiency) were duplicated. The mercury-arc tube requires a starting electrode and usually a mechanical tilting device for starting.

18. Igniter-type Mercury Rectifier. A mercury-pool cathode provided with a thyatron igniter is called the *ignitron*.² This rectifier is characterized by an

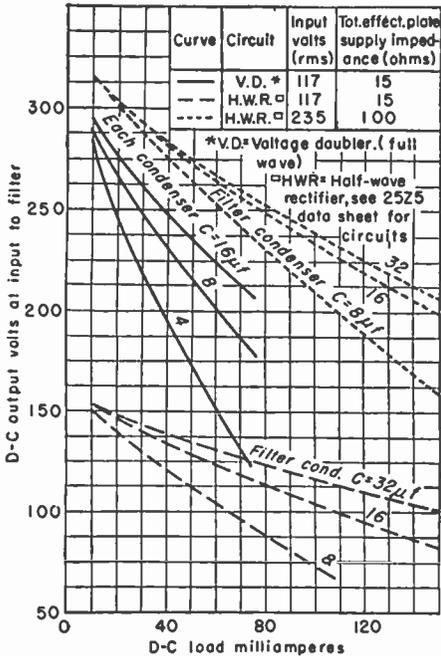


FIG. 18. Operation characteristics, 25Z6 rectifier.

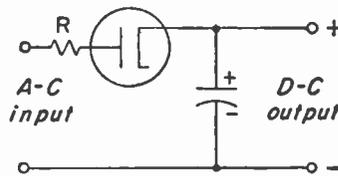


FIG. 19. Half-wave rectifier.

efficiency of about 90 per cent for loads of 10 to 125 per cent, for a typical 300-kw 275-volt d-c output from a 2,300-volt, three-phase, 60-cycle transformer. This is higher efficiency over a wider range of loads than is possible from a synchronous converter or a synchronous motor-generator set. Ignitrons are particularly useful for supplying high current below 600 volts and are not usually employed for radio applications.

¹ Prince, D. C., and F. B. Vogdes, "Principles of Mercury Arc Rectifiers and Their Circuits," p. 23. McGraw-Hill, 1927.

² Pronounced "ig-nye-tron."

19. Dry Metallic Rectifiers.¹ At present, metallic rectifiers available for general use are of three types: magnesium copper-sulfide,² copper oxide,³ and selenium.⁴

20. Copper-sulfide Rectifier. This rectifier is assembled from disks of cupric sulfide in intimate contact with magnesium washers, with or without radiating fins and spacers, mounted on bolts, and clamped together under high pressure.

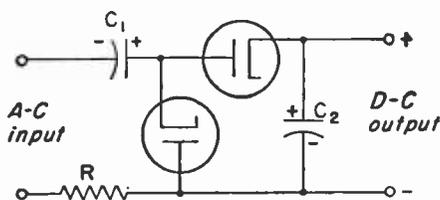


FIG. 20. Half-wave doubler.

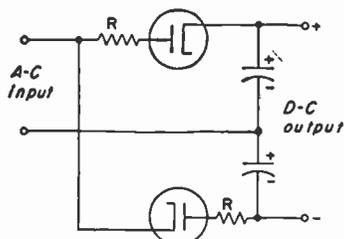


FIG. 21. Full-wave doubler.

The rectifier is characterized by small size and weight, low initial cost, good voltage regulation, and relatively low efficiency. Normal operating temperature is high, limited to 130°C max, although temperatures above this may be permitted in special cases. The rectifier has a definitely limited life.

Overloads in current are allowed for intermittent duty up to four or five times normal rating, provided the maximum temperature limit is not exceeded.

Individual disks have a low voltage rating, the maximum reverse voltage per disk being 4 volts a.c. rms. Current ratings are quite high, reaching 40 amp per sq in. for forced-draft ventilation.

There is some difficulty in operating rectifier elements in parallel, generally requiring separate transformer secondaries for each rectifying element. There is also some limit to the number of elements that can be operated in series, so that most applications¹ are made in the range below 50 volts. The rectifier is not used in very low-power work, such as in instruments.

Most suitable applications for the copper-sulfide rectifier appear to be for intermittent duty, where the life can be stretched out over a longer time, where efficiency is not a factor, and where in addition the first cost is important.

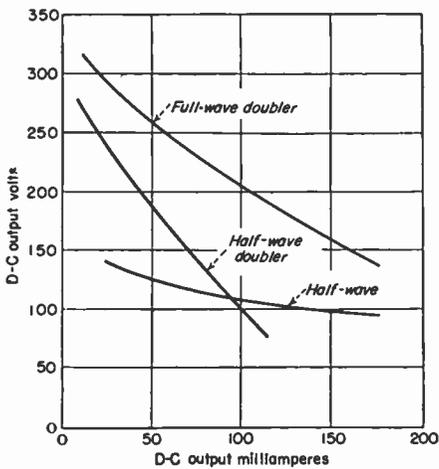


FIG. 22. Comparative voltage regulation, rectifier circuits.

Among such applications are the operation of various types of electromagnetic loads, such as circuit-breaker solenoids. Other applications include home battery chargers, quick chargers, engine starters, and plating rectifiers.

21. Copper-oxide Rectifiers. This rectifier is built up from copper disks oxidized on one side, or from plates, which may be oxidized on both sides. In the disk-type

¹ Sections 19 through 24, and Figs. 23 and 24, were furnished by I. R. Smith, Motor Engineering Department, Westinghouse Electric Corp., East Pittsburgh, Pa.

² B-L Electric Corp.; P. R. Mallory Co.

³ General Electric Co.; Westinghouse Electric Corp.

⁴ Federal Telephone and Radio Corp.; General Electric Co.; Westinghouse Electric Corp.; Selenium Corp of America; Fansteel Metallurgical Corp.; B-L Electric Corp.; Radio Receptor Corp.; Bradley Laboratories.

assembly, lead washers are placed next to the oxide surface to improve the contact, radiating fins and spacers are included when necessary, and the whole assembly, usually mounted on an insulated bolt, is clamped together under high pressure. Such stacks are generally for use with natural ventilation, with output ratings of 4.5 volts d.c. per element in a single-phase bridge, with current ratings up to 0.4 amp per sq in. Forced ventilation may be used to increase these ratings to 5 volts d.c. and 0.8 amp per sq in.

“High voltage” units are available with ratings under forced draft up to 20 volts d.c. per element in a three-phase bridge.

Plate-type assemblies are normally used with forced draft and for larger power outputs. Output ratings of 6 volts d.c. per element from a three-phase bridge are normal, with current ratings of 0.6 amp per sq in. Plates are spaced $\frac{1}{4}$ to $\frac{1}{2}$ in. apart. Contact to the plate surface is made with spring washers at light pressure, the outer surface of the oxide having been electroplated with nickel for the negative electrode. When used with natural ventilation, three-phase ratings reduce to 4.8 volts d.c. per element at 0.275 amp per sq in.

Disk sizes available range from 0.090 to 1.5 in. in diameter. Plate sizes range from $4\frac{3}{8}$ in. square to $4\frac{3}{8}$ by 12 in., the latter having 62 sq in. of active area.

The copper-oxide rectifier is characterized by its indefinitely long life, operating records approaching 20 years of service having been established. The rectifier has good efficiency and high thermal capacity. Voltage regulation is relatively poor. Size and weight for a given rating are larger than for the other types. Initial cost is higher than for copper sulfide.

Temperature rise in operation is generally kept below 15°C. Standard ratings are based on 35°C max ambient, derating in voltage being required for higher ambients.

Figure 23 shows typical forward-direction volt-ampere curves for a standard copper-oxide unit, together with those for copper sulfide and selenium. Back emf is specified in Secs. 20 to 22.

Applications for this rectifier are almost unlimited. It is best fitted for those purposes which require long life and good efficiency and where the first cost is less important than the operating cost. High thermal capacity and ability to withstand high short-time overloads in voltage and current have made the rectifier very useful in intermittent-duty work, especially the operation of circuit-breaker solenoids, airplane-engine starters, etc.

Other typical applications include instrument rectifiers delivering 0.001 watt, and plating and welding rectifiers delivering as much as 120 kw from a single cubicle. Important uses are battery charging, including quick chargers, industrial truck and railway application, cathodic protection, electroplating, cleaning, and anodizing, operation of all kinds of d-c control—contactors, brakes, etc.—operation of other magnetic devices such as separators and lifting magnets, dynamic braking, operating of business machines, and many others.

22. Selenium Rectifiers. This rectifier is assembled from aluminum or steel disks or plates coated with selenium, which in turn is coated with a sprayed alloy. Commonly each unit is assembled with a spring washer which contacts the sprayed metal surface under relatively light pressure. The disks and spring washers with any required spacers or cooling fins are assembled on insulated bolts and clamped together.

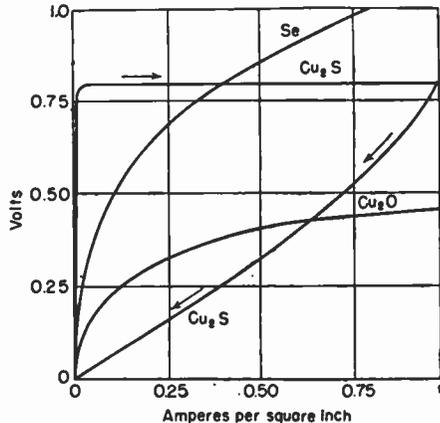


FIG. 23. Comparative volt-ampere characteristics of metallic rectifiers.

A second method of assembly avoids the use of individual spring washers, contact being made under relatively high pressure to a small nonrectifying area around the center hole.

Standard lines of selenium rectifiers include six or seven sizes, ranging from 1 to $4\frac{3}{8}$ in. in diameter, plus larger rectangular plates up to approximately 5 by 7 in. These may be used either with natural or forced ventilation.

Ratings are based on 35°C max ambient temperature, units being derated in voltage above 50°C and in current above 35°C.

Standard units are rated at 18 volts a.c. rms maximum reverse voltage, corresponding to an output voltage from a single-phase bridge of 12 volts d.c. per disk. "High voltage" elements are also available, the d-c voltage obtainable from a single-phase bridge being about 20. Still higher voltage elements are being used in radio applications to replace rectifier tubes. Such cells have a reverse voltage rating of 72 volts peak per element.

Current ratings for natural ventilation are based on 0.32 amp per sq in. with some manufacturers permitting an increase up to 0.48 amp per sq in. for short stacks. When forced draft is used, current ratings are increased up to 0.8 amp per sq in., the voltage ratings remaining unchanged.

For intermittent duty, overloads are permitted in current but not in voltage. Some care has to be exercised in applications on highly intermittent operation because of the possibility of unforming the rectifier barrier layer.

The selenium rectifier is notable for its small size and weight for a given wattage output, especially as compared to the copper-oxide and somewhat less so when compared with the copper-sulfide rectifiers. Efficiency is comparable to that of copper oxide, while voltage regulation is better, being usually not over 10 per cent.

Applications in general are the same as listed for copper oxide, except that the rectifier is generally not considered as suitable for instrument use, where copper oxide is preferred.

23. Circuit Connections. Most dry metallic rectifier applications can be handled with one of the seven simple circuits shown in Fig. 24. Most common are the single-phase and three-phase bridge circuits, which give best transformer utilization and can in fact be used without transformers when the voltage requirements are suitable. The three-phase bridge circuit has the advantage that the d-c output voltage is roughly the same as the a-c input voltage.

Center-tap connections provide points of maximum efficiency at voltages halfway between the points of maximum efficiency obtainable from a bridge.

The half-wave circuit is used with a capacitor filter to supply light loads, or where the output current required is less than half of the rating of the element in a bridge. Half-wave and center-tap circuits are both used when the bridge connection would work the unit at less than half its voltage rating, thus reducing the size of the rectifier and improving the efficiency.

Voltage-doubler circuits are used to supply low currents at relatively high voltages.

24. General Conclusions on Dry Metallic Rectifiers. A brief comparison of the three rectifiers shows the copper-sulfide rectifier to be lowest in first cost, life, and efficiency, and best in regulation. The selenium rectifier has the smallest size and weight for a given rating, better efficiency, and longer life than the copper-sulfide. The copper-oxide rectifier has the longest established life, is equal in efficiency to the selenium, is best for instrument use and probably for intermittent loads, and is the only one entirely free from unforming effects during idle periods.

25. Low-power Transformers for Receivers and Amplifiers. The design of a reliable power transformer having high efficiency requires fairly elaborate calculations. To take into account the d.c. which flows in a transformer secondary when a half-wave rectifier is used, some interesting equations have been derived.¹

A simple approximate-design method will be given for the construction of a single-phase low-powered transformer up to 180 volt-amp; 180 watts if power factor is near

¹ Harder, E. L., *Electric J.*, October, 1930, p. 601.

unity. This design applies to transformers which supply full-wave rectifiers, and energy to a-c powered radio receivers, or power amplifiers.

26. **Small Transformer Details.** Economy in a transformer is secured when the winding encloses a maximum of core area with a minimum of wire and when the magnetic path is as short as possible

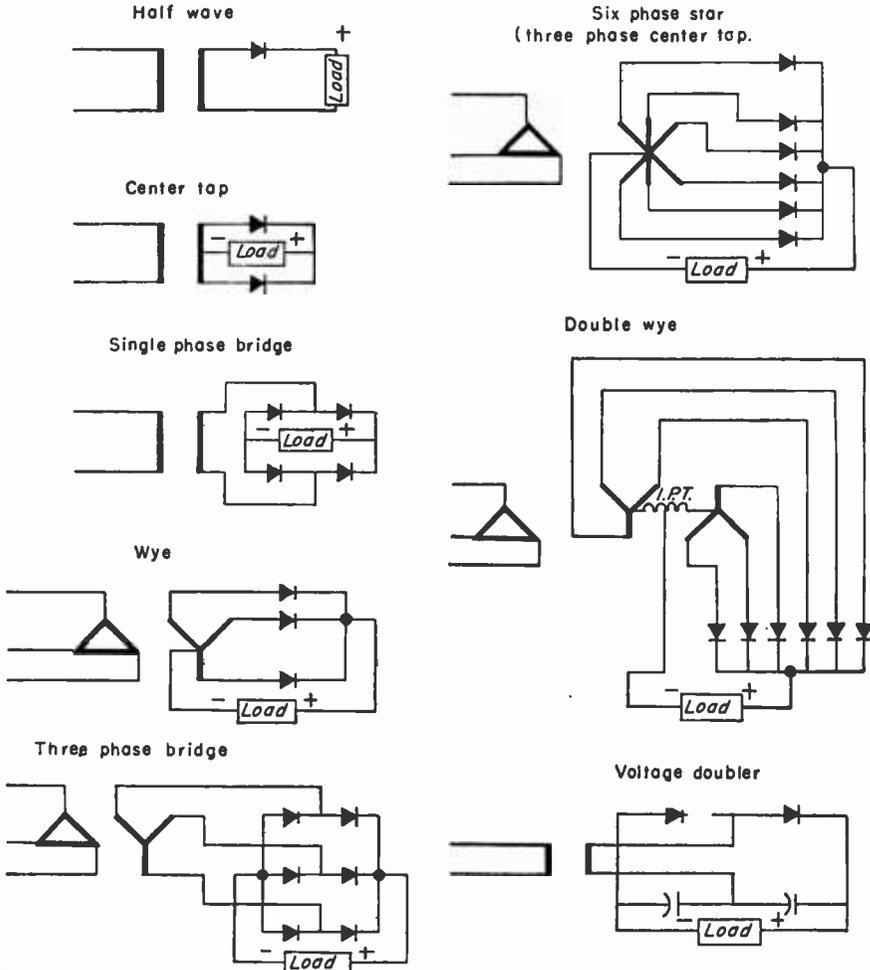


FIG. 24. Circuits employed with dry metallic rectifiers.

The core form of a small transformer can be of several stock shapes, one popular punching being in the form of a capital letter E. As a rule, two punchings are employed, one having longer legs than the other, so that the magnetic circuit breaks joints in stacking the iron. Another convention usually followed in small transformers is the use of a single winding form, all secondaries and primary being on the middle leg of the E core (see Fig. 25).

The winding form, or spool, is usually an insulating tube to which side pieces of

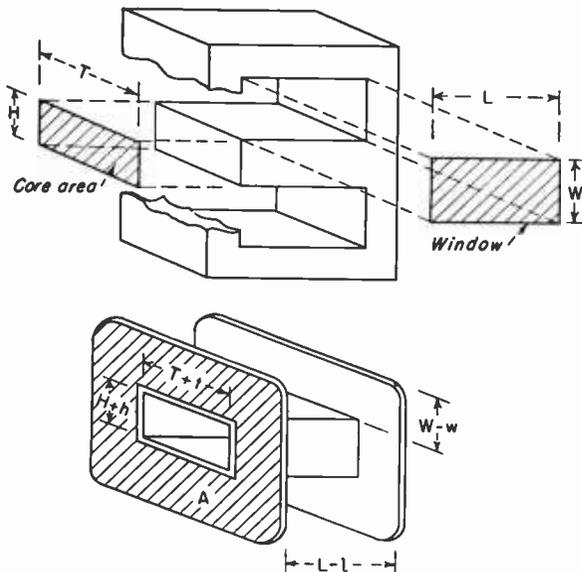


FIG. 25. Transformer core and spool. Volume of coil = $(L - l)A$.

insulation are fitted and on which terminals are placed. If the coil is to be machine-wound with interwoven insulation, the side pieces are often omitted in favor of flexible leads.

27. Ten Steps in Designing a Small Power Transformer.

1. Determine the Volts and Amperes Needed for Each Secondary.

- a. Total maximum secondary watts = $W_s = E_1I_1 + E_2I_2 + E_3I_3 + \dots$
- b. Total watts needed for primary = W_p . Assuming 90 per cent efficiency, $W_p = W_s/0.9$.
- c. Find primary amperes assuming 90 per cent power factor,

$$I_p = \frac{W_p}{E_p \times 0.9} = \frac{W_s}{0.81E_p}$$

and for primary $E_p = 115$ volts, $I_p = W_s/93.1$ amp.

2. Size of Wire. Knowing the current for each winding, the wire size is determined by the circular mils per ampere which it is desired to use. A rule of thumb is 1,000 cir mils per amp for each secondary winding, and 1,500 cir mils per amp for the primary.

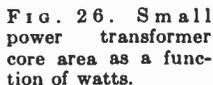


FIG. 26. Small power transformer core area as a function of watts.

3. Core Considerations. A curve of core areas for different powers and 60 cycles is Fig. 26. For 40 watts the area is 1 sq in.; 70 watts, 1.5 sq in.; 120 watts, 2 sq in. The core area is also shown in Fig. 25. The area of the core is the same as the inside dimensions of the spool after making a 10 per cent allowance for stacking. For example, a spool 1 by 2 in. inside would enclose 2 sq in. but allowing for a 10 per cent loss, only 90 per cent or $0.9 \times 2 = 1.8$ sq in. would be the net core area. The net core area is abbreviated to "core area" when determining turns per volt.

The upper part of Fig. 25 shows the core area $T \times H$ and the window or winding space as $L \times W$. The spool must be slightly different to allow ready assembly, since

the core area $T \times H$ is surrounded by the spool whose dimensions are $(T + t) (H + h)$. The lower-case letters indicate *small changes* in the dimensions. Note that the spool dimensions to fit the window are *slightly smaller* than $L \times W$, being $(L - l) \times (W - w)$.

4. *Core Loss and Induction.* The flux density at which the core is to be worked determines the iron (core) loss. Figure 27 gives several different core materials, watts per pound at 60 cycles being plotted against flux densities (see Table 3 for higher frequencies). An average working value for the induction is 65 kilolines per sq in. or 10 kilogausses. Curves shown in the figure depend mainly on experimental data, not directly on theory. For this reason no definite value of core loss for every

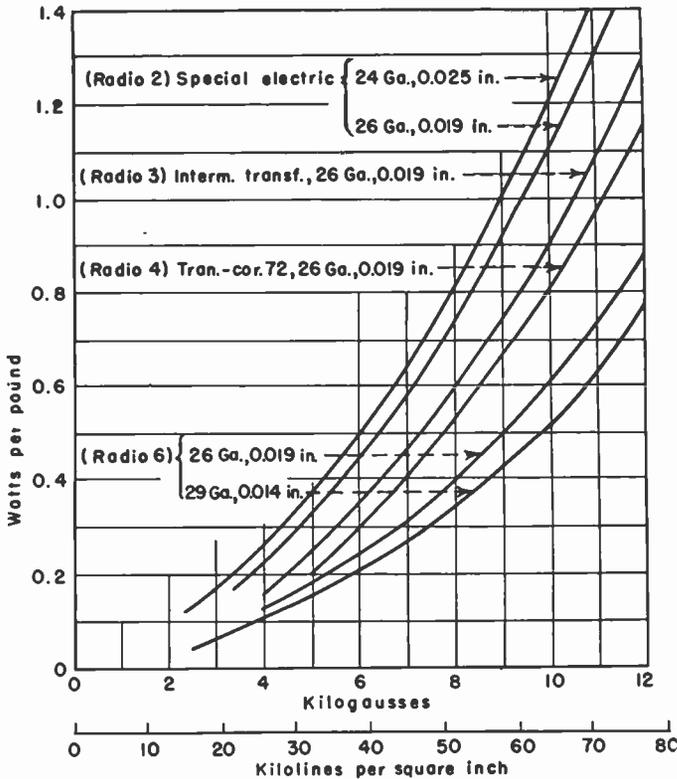


FIG. 27. Core-loss curves for Armeo radio grades.

magnetic circuit can be stated; such losses depend on the quality of core material available. It can be said that better and better core material is constantly being made, having lower loss per pound, so that the use of higher flux densities is becoming possible. Induction up to 15 kilogausses is not uncommon, though unusual for a transformer. The core loss increases with frequency, a typical curve being Fig. 28.

As indicative of newer core materials, that with "flux parallel to the rolling direction" may be mentioned.¹ This material has a permeability much higher than the magnetic steels shown on Fig. 27. It has a permeability of 56,000 at 6,000 gausses, 25,650 at 10,000 gausses, and 2,090 at 14,000 gausses, for a sheet 0.013 in. in thickness, at 60 cycles. The typical 60-cycle core loss for this material for the flux densities

¹ Electrical Alloy No. 48 (Armeo); Hipersil (Westinghouse Electric Corp.)

mentioned are, respectively, 0.052, 0.132, and 0.255 watt per lb. This core material is fully annealed, after shearing, by being held at 2200°F for 30 hr in pure dry hydrogen. It is cemented into a solid mass, then cut apart for the insertion of the coil. Straps hold the core together in place around the coil, in the final assembly.

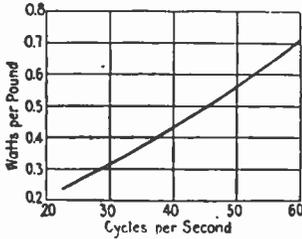


FIG. 28. Core loss vs. frequency, 29-gage Armco Trans-Core, for an induction of 10,000 gausses.

Since high-temperature-stable insulations for wire are becoming more common, it is worth noting that this core material is not expected to change appreciably at temperatures up to 150°C; *i.e.*, there is no magnetic aging effect. It is not expected that the core will heat itself up to this high value, but it is possible for a high copper loss to do so by its heat generation.¹

Table 3 shows some characteristics of thin-gage electrical steel for frequencies higher than 60 cycles, starting with airplane 400-cycle transformers,² and going up to 100,000 cycles.

Table 3. Maximum Values of Core Loss for Armco Thin-gage Electrical Steel

Thickness, in.	Frequency, cycles	Induction, kilo-gausses	Grain of steel	Core loss, watts/lb
0.002	100,000	0.25	Parallel	5*
0.003	10,000	1.00	Parallel	6
0.005	1,000	5.00	50-50	7
0.007	400	10.00	50-50	8

* Typical maximum value.

5. *Induced-voltage Equation. Turns per Volt.* The fundamental definition that a potential difference of 1 volt is induced when 10^8 magnetic lines (maxwells) are cut in 1 sec is the basis of the equation

$$E = \frac{BANf}{10^8} \times 4.44 \quad (1)$$

where E = volts

A = area of core

B = flux density, whose denominator is same units as A

f = cycles

N = number of turns

A more useful working equation for small power transformers is obtained by rewriting the equation to give turns per volt, N/E ,

$$\frac{N}{E} = \frac{10^8}{BAf4.44} \quad (2)$$

Figure 31 is an alignment chart of this equation. Equations (1) and (2) can readily be used for frequencies other than 25 and 60, though the chart of Fig. 31 is restricted to these frequencies.

Using a flux density of 65 kilolines per sq in. and the net core area mentioned in step 3 (1.8 sq in.), N/E for 60 cycles is found to be 3.1, from Fig. 31. Thus for each

¹ A great deal of careful work in supplying magnetic data, notably Figs. 27, 45, 47, Table 3, and other information used in Secs. 27 (step 4) and 38, was furnished by Mr. G. H. Cole, Associate Director, Research Laboratories, The American Rolling Mill Company, Middletown, Ohio.

² Detailed analysis of transformers, for 100 to 3,000 cycles can be found in "Principles for the Design of Small Power Transformers," *Harvard Univ. Radio Res. Lab. Rpt.* 411-478, January, 1944. Photostat obtainable from the Library of Congress, Washington, D.C.

volt supplied to or obtained from the transformer, there must be 3.1 turns of wire. It is permissible to use decimal fractions in the use of turns per volt; often this is advisable, to get the maximum performance out of a minimum of material. Further, a change of a few per cent in N/E will change the induction very slightly, and since this is an approximate method, the change will probably not be noticed. To get the fractional voltages needed, say for a 6.3-volt heater, some additional turns are usually provided for IR drop, so the use of a fractional N/E is no hardship.

The voltage drop in the transformer secondary windings will be mentioned here in general, and again in the specific example, Sec. 29. The voltage available at a tube heater is less than that at the terminals of the transformer secondary, and the voltage

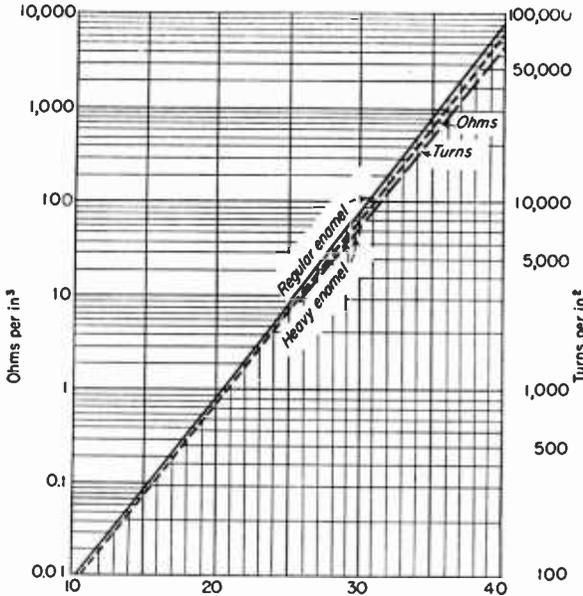


FIG. 29. B & S gage round copper wire for theoretical coils. For actual coils, allow for space factor.

at the terminals of the secondary will also be lower than that induced by the magnetic lines in the core. This is due partly to the IR drop in the secondary winding and partly to the drop in the connecting wires to the heater. The IR drop in the secondary will be calculated, and the IR of the connecting wires will be allowed for by adding a few extra turns on the secondary.

6. *Turns for Each Winding.* In step 1 the desired voltages were given, E_1 , E_2 , etc. Using the turns per volt found in step 5, the total turns for each winding are found; e.g., with 3.1 turns per volt, a 117-volt winding should have $3.1 \times 117 = 316$ turns.

7. *Winding Space Required.* From the total turns for each winding and the wire size, the total area of winding space is calculated. Different wires and insulations have definite (maximum) turns per square inch. The method of insulation and winding may change these maximum values by a factor as much as 0.3 (for the actual winding) to 1.0 for the maximum. Specifically, a 900-turn coil of regular enamel wire of B & S No. 20 copper wire can, theoretically, be wound in 1 sq in. of space. This is the *maximum* number of turns. If thin insulating paper is interleaved between layers, the area remaining for the wire is less. If cotton is interwoven back and forth

over the coil as the turns are being put on, the final coil may show an over-all winding space of half the maximum turns per square inch. Figures 29 and 30 give data on coil windings.

Figure 29 shows both turns per square inch and ohms per cubic inch for round copper wires B & S Nos. 10 to 40, for regular and heavy enamel.¹ In Fig. 29 there is no allowance for interwoven cotton, insulation between layers, or loss of space due to wires not being laid tightly together. It is the theoretical maximum both of turns and of ohms. For example, for regular enamel B & S No. 30, a maximum of 8,500 turns will go into each square inch of window space and will have a maximum of 80 ohms per cu in. of the winding.

Though calculated for enamel copper wires, Fig. 29 is readily used for other coverings, if the relative space factor is known. As a first example, suppose No. 30 DCC copper wire is to be used. From Fig. 30 the relative space factor is 40 per cent, thus

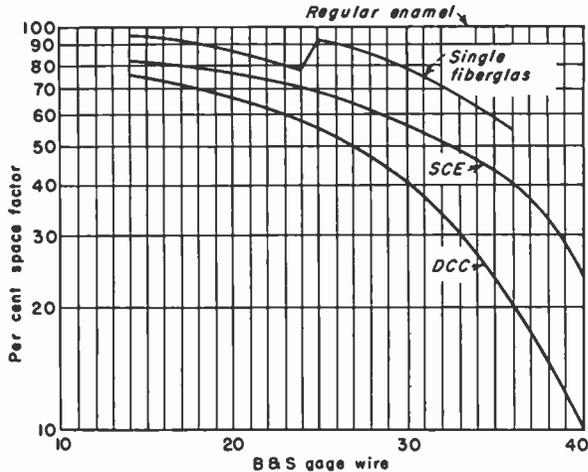


FIG. 30. Per cent space factor for round insulated wire, compared to regular enamel.

the maximum turns per square inch for this new wire will be $0.40 \times 8,500 = 3,400$ turns per sq in., and the maximum resistance is $0.40 \times 80 = 32.0$ ohms per cu in. of winding.

As a second example, suppose that cotton is interwoven in the coil so that the space factor is 70 per cent due to the cotton alone. This will mean for regular enamel copper wire a winding of $0.70 \times 8,500 = 5,950$ actual turns per sq in., with a resistance of $0.70 \times 80 = 56$ ohms per cu in. However if the No. 30 DCC wire is used, the resulting values are $0.40 \times 0.70 \times 8,500 = 2,380$ turns per sq in. and $0.40 \times 0.70 \times 80 = 22.4$ ohms per cu in.

Note that the solid line on Fig. 29 serves to show both turns and ohms for regular enamel copper wire. For heavy enamel the dotted line is for ohms, and the dashed line is for turns. Thus for No. 30 heavy enamel copper wire the maximum turns per square inch is 7,400, and the ohms per cubic inch is 70.

Both Figs. 29 and 30 are intended for approximate preliminary calculations. Actual space factors are difficult to predict exactly. Sample coils should be wound and checked before large-scale production is authorized.

After the winding space of all the windings is added up, the area is compared with that of the core window. If the winding will go in the space, this part of the design is finished. If the wires will not go in the space available, there are several approaches

¹ Data for Figs. 29 and 30 were taken from catalogue of Belden Manufacturing Company, Chicago.

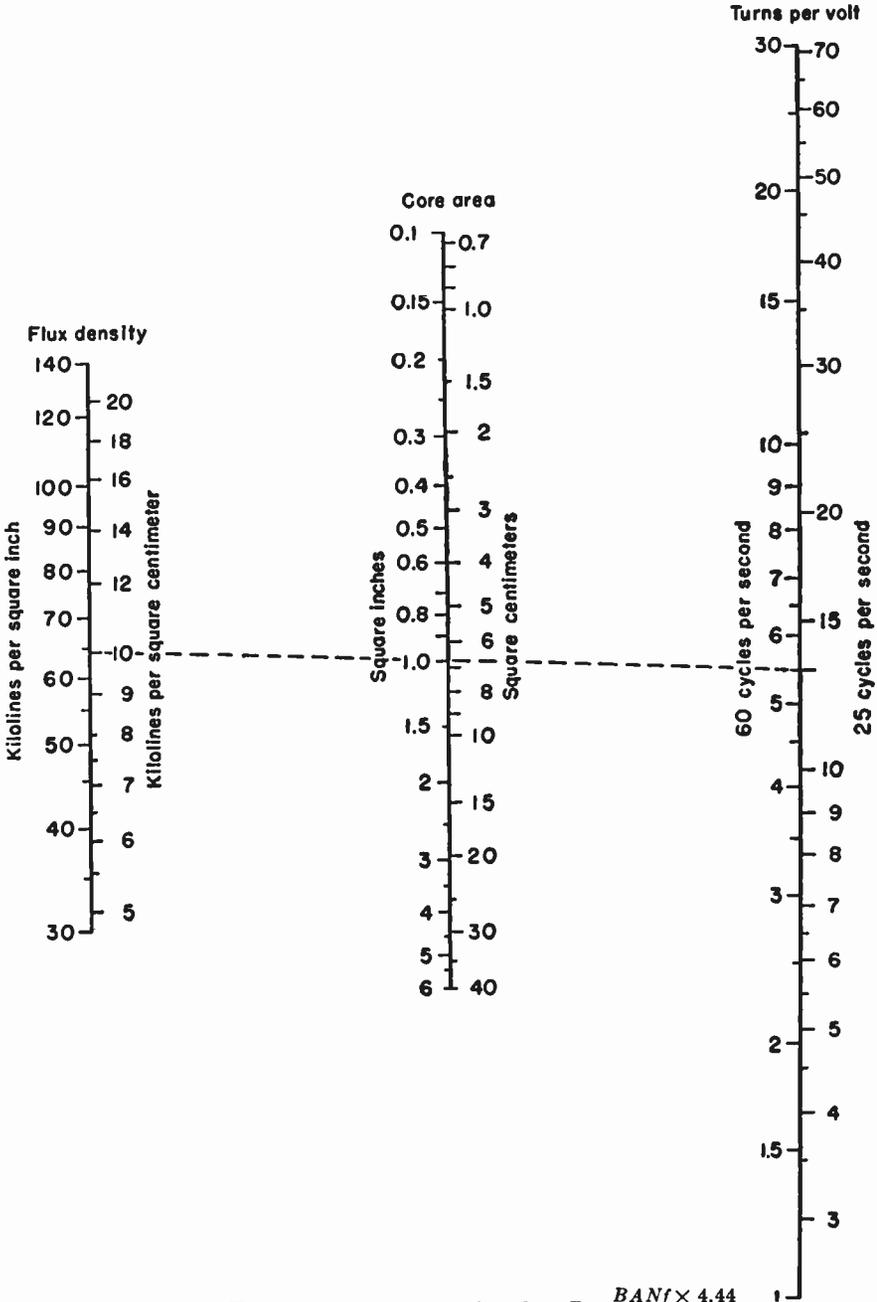


FIG. 31. Transformer design chart based on $E = \frac{BANf \times 4.44}{10^8}$

to a solution. From the standpoint of the coil, the wires may be smaller in diameter, the insulation may be thinner, or fewer turns may be used (increase the induction of the iron). From the standpoint of the iron, a thicker stack can be used, which, by increasing the core area will require fewer turns per volt for the same induction, thus decreasing the cross section of the winding.

8. *Copper Loss.* Two methods are available: (a) Since Fig. 29 gives ohms per cubic inches and the area in square inches can be computed to see if the winding will go in the space, it only requires a knowledge of the average (or mean) turn to get cubic inches, and then the resistance of each winding. For example, if the mean turn is 10 in. long, a 1,190-turn coil of No. 30 regular enamel, space factor 0.70, has an area of 0.2 sq in. The volume of the coil is $0.2 \times 10 = 2.0$ cu in. Then recalling from Sec. 7 that the resistance is 56 ohms per cu in., we get the resistance of the coil, $2.0 \times 56 = 112$ ohms.

(b) An alternative method uses wire tables: Find the length of the mean turn in feet; find the length of each winding in feet by multiplying the number of turns by the mean-turn length; from wire tables find the ohms per 1,000 ft for this size wire, and then the ohms for the actual length.

(c) For either method (a) or (b): Multiply the current squared, for each winding, by the ohms for that winding; add the I^2R s for each winding to get copper loss L_1 .

9. *Core Loss.* The core loss in watts L_2 is found from (a) the weight of the core, (b) the flux density, and (c) the kind of core used in step 4. A useful factor is that 4 per cent silicon steel weighs 0.27 lb per cu in. of volume.

10. The approximate percentage efficiency is $\frac{W_s \times 100}{W_s + L_1 + L_2}$, W_s being the secondary watts (see step 1).

NOTE. If step 10 shows about 90 per cent efficiency, the preliminary design may be regarded as satisfactory. If much less than 90 per cent, step 1(a) must be modified, a new larger value of I_p is used, and a larger primary wire size employed. This will increase the efficiency since L_1 will now be smaller. However, additional winding space is needed, and a check on the available window area must be made.

28. *RMA Transformer Color Code.* When flexible leads are brought out from a power transformer, the approved color coding for the insulation is as follows:

Leads	Wire Insulation Color
1. Primary lead start	Black
Tap (if used)	Black and yellow stripe (50-50*)
Primary finish	Black and red stripe (50-50)
2. High-voltage secondary winding (start and finish)	Red
Center tap for high-voltage secondary	Red and yellow stripe (50-50)
3. Rectifier filament winding (start and finish)	Yellow
Center tap on filament winding	Yellow and blue stripe (50-50)
4. Filament winding No. 1	Green
Center-tap filament winding No. 1	Green and yellow stripe (50-50)
5. Filament winding No. 2	Brown
Center tap on No. 2	Brown and yellow stripe (50-50)
6. Filament winding No. 3	Slate
Center tap on No. 3	Slate and yellow stripe (50-50)
* That is, two stripes of equal width	

29. *Typical Small Transformer Design.* This transformer is intended to work from a 115-volt 60-cycle source and to furnish power for A , heater; B , plate; and C , bias for the power amplifier circuit of Fig. 9. The required output from the filter to the amplifier is 425 volts d.c. Allowing for two filter chokes L_1 and L_2 of 100 ohms each in the circuit of Fig. 35, and a d-c load of 200 ma, the rectifier input must be $425 + 0.2 \times 200 = 465$ volts d.c. The rectifier tube will be a 5Z3 or 5U4. Figure 17 shows that a transformer input of 450 volts rms per plate is required. This is sometimes written as 450-0-450 indicating a center tap by the use of the 0.

The rms current for a full-wave rectifier to supply 200 ma d.c. is $200 \times 1.11 = 222$ ma, as noted in Sec. 3 in connection with Fig. 1.

1. The desired transformer secondary volts and amperes are as follows:

Volts, <i>E</i>	Amp, <i>I</i>	Purpose	Watts, <i>EI</i>
450	0.111	<i>B</i> and <i>C</i> supply	50.0
450	0.111	<i>B</i> and <i>C</i> supply	50.0
5.0	3.0	Rectifier filament	15.0
6.3	2.4	Heaters $(2 \times 0.9) + (2 \times 0.3)$	15.1

a. Total secondary watts $W_s = 130.1$.

b. Primary watts $W_p = W_s/0.9 = 130.1/0.9 = 145$ watts.

c. Primary amperes $I_p = W_s/93.1 = 130.1/93.1 = 1.40$ amp (from 115 volts).

2. The secondaries will use 1,000 cir mils per amp and the primary 1,500 cir mils per amp; the wire sizes and the identifying amperes and volts are listed in the accompanying table. The use of larger wires keeps down the *IR* drop and the I^2R losses, but larger wires use more space and may have to be changed if the winding cross section is too large.

Volts	Amp	Cir mils	Nominal size of wire, B & S (from wire table)	Actual size* of wire, B & S
115	1.40	2,100	17	
450	0.111	110	29	
5.0	3.0	3,000	15	
6.3	2.4	2,400	16	

* This column is left blank for the initial design. It may be filled in later (1) if the transformer coil does not fit the space allotted, in which case the wire sizes may be revised, or (2) if the use of several sizes of wire is objectionable. As shown in the table above, B & S wire sizes 15, 16, and 17 are called for. These might be changed to read No. 16 for each, since heater windings have negligible *IR* drops, as will be shown in step 8, and the primary might, to advantage, be wound with a larger wire as it has the greatest I^2R loss of any winding (step 8).

3. The core area of the 145-watt transformer, from Fig. 26, is 2.2 sq in. To make this net core area with 90 per cent stacking factor requires an actual core area of $2.2/0.9 = 2.45$ sq in. so $T \times H = 2.45$ sq in. on Fig. 25. The center leg of the punching available has $H = 1\frac{3}{16}$ in.; thus the stack must be $2.45/1.19 = 2.06$ in. thick = *T*.

4. The flux density to be used is 65 kilolines per sq in. (10 kilogausses) and the core is 4 per cent silicon steel having a 60-cycle loss of 0.6 watt per lb of weight. The core 2.06 in. thick weighs 5.15 lb, so the iron loss $L_2 = 5.15 \times 0.6 = 3.09$ watts.

5. For 65 kilolines per sq in. and core area 2.2 sq in., the turns per volt, from Fig. 25, are found to be 2.6 for 60 cycles.

6. The turns for each winding are now calculated (table at right):

7. The winding space, in square inches, using regular enamel wire, is taken from Fig. 29 and tabulated (column *C*, below), and the minimum space *D* equals the quotient of columns *A* and *C*.

Volts	Nominal turns	Actual turns*
115	299	300
450	1,170	1,200
450	1,170	1,200
5.0	13	14
6.3	16.4	18

* For primary and high-voltage windings, a change of a few per cent (up) is justified at this preliminary stage. For the heavy current secondaries, an extra turn or two is added to compensate the *IR* drop in wires and soldered junctions.

<i>A</i>	<i>B</i>	<i>C</i>	<i>D = A ÷ C</i>	<i>E</i>
Turns	B & S	From Fig. 29, turns per sq in. regular enamel	Min space, sq in.	Allowable space when revised,* sq in.
300	17	450	0.67	1.20
1,200	29	6,700	0.18	0.32
1,200	29	6,700	0.18	0.32
14	15	290	0.05	0.09
18	16	370	0.05	0.09
Total.....	1.13	2.02

* Column *D* is written down first and compared with the actual window winding space, which here is found by using $L = 2.0$ in., $W = 1.0$ in.; $L \times W = 2.0$ sq in. Since *D* is 1.13 sq in. and allowable space is 2.0 sq in., the space factor is $1.13/2.0 = 0.56$, and the space of each item in column *D* is divided by this factor, giving column *E*, the actual space that can be allowed each winding. Note that the heavy current secondaries take very little space, and a revision of their wire sizes would cause little change in the required winding space. The 56 per cent space factor is satisfactory for a preliminary shop trial of this coil design.

8. The copper loss is found by calculating the I^2R loss for each winding and adding them. This may be done rapidly by using Fig. 29, data already found, and knowing the mean turn $l = 11$ in. [see (a) below]; or more accurately by wire tables and using $1\frac{1}{2}$ ft for the mean turn [see (b)].

(a) From Fig. 29 and data previously found (note that the minimum space D from the previous table is used for this calculation, rather than the allowed space E):

A	D	$F = D \times l$	G	$H = G \times F$	IR	I^2R
Turns	Min space, sq in.	Volume, cu in.	Ohms per cu in. (Fig. 29)	Ohms	Volts drop	Watts loss
300	0.67	7.5	0.2	1.5	2.94
1,200	0.18	2.0	45	90	1.09
1,200	0.18	2.0	45	90	1.09
14	0.05	0.55	0.08	0.04	0.12	0.36
18	0.05	0.55	0.12	0.07	0.17	0.40
Total.....	5.88

By this method the copper loss $L_1 = 5.88$ watts.

(b) From wire tables and knowing that the mean turn is $1\frac{1}{2}$ ft = l :

A	$J = A \times 1\frac{1}{2}$	K	$H = J \times K/1,000$	IR	I^2R
Turns	Feet	Ohms per 1,000 ft (wire table)	Actual ohms	Volts drop	Watts loss
300	275	5.16	1.42	2.78
1,200	1,100	83.4	92	1.11
1,200	1,100	83.4	92	1.11
14	12.8	3.25	0.04	0.12	0.36
18	16.5	4.09	0.07	0.17	0.40
Total.....	5.76

From this method L_2 the copper loss is 5.76 watts, differing from the quicker method of (a) by less than 2 per cent.

9. Total watts output = 130.1 (step 1).

Losses $L_1 + L_2 = 5.9 + 3.1 = 9.0$ (steps 4 and 8).

$$\text{Efficiency} = \frac{130.1 \times 100}{130.1 + 9.0} = \frac{13,010}{139.1} = 93 \text{ per cent.}$$

10. The volts drop are not serious as shown by the small IR in step 8.

30. Filters for Smoothing Rectified and Generated D.C. To supply d.c. from rectified a.c. requires a low-pass filter.¹ Low-pass filters are divided into two classes: tuned and untuned. The tuned filter offers a maximum impedance or attenuation to the frequency of the supply, but the impedance at nearby lower or higher frequencies is not quite so great (Fig. 32), although the general trend of the curve is a rising attenuation as the frequency increases.

A common form of untuned low-pass filter is shown in Fig. 35, using three capacitors and two inductors (chokes). This filter has a continuously rising curve of impedance

¹ For filter theory, see Johnson and Shea, "Transmission Circuits for Telephone Circuits," Van Nostrand, 1925, and G. W. Pierce, "Electric Oscillations and Electric Waves," McGraw-Hill, 1920.

as the frequency increases. To get good results with this filter, it is desirable to choose f_c , the frequency at which attenuation begins, as low as possible. The simplest equations for determining the inductance and capacitance for this filter are as follows:

$$C = \frac{1}{\pi f_c R} = \frac{0.3183}{f_c R} \tag{3}$$

$$L = \frac{R}{\pi f_c} = \frac{0.3183R}{f_c} \tag{4}$$

where f_c = frequency where attenuation begins, cycles
 C = capacitance, farads
 R = load resistance, ohms
 L = inductance, henrys

As these equations are often used as preliminary steps in designing filters, Fig. 33 has been made up to give the data in convenient chart form. With two of the factors

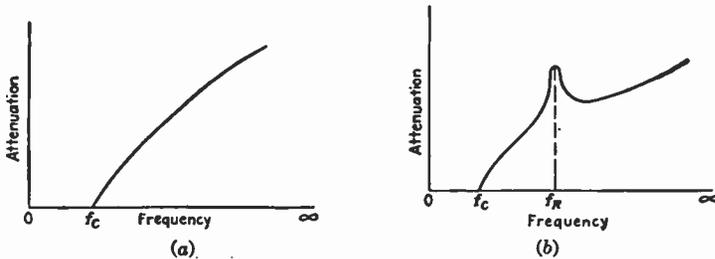


FIG. 32. (a) Low-pass filter. (b) Tuned low-pass filter.

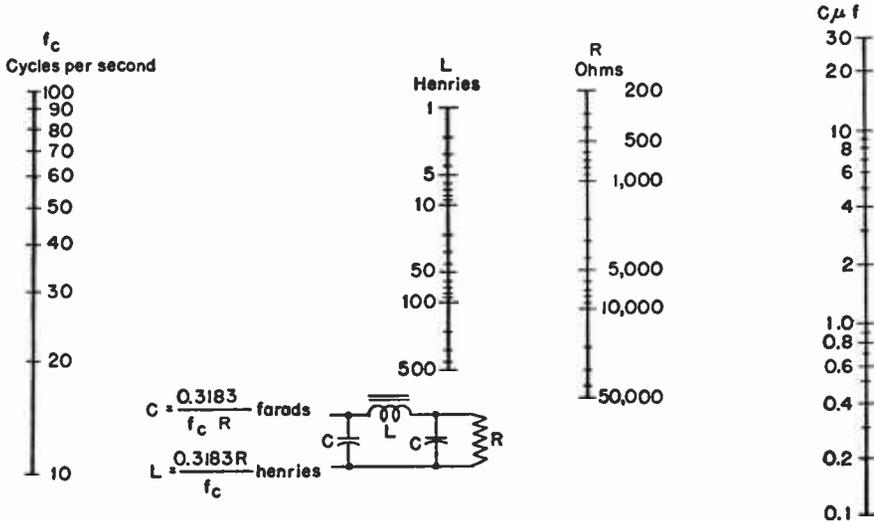


FIG. 33. Low-pass filter design chart, pi section.

known, the other two quantities are determined from the chart by a straightedge across the two known quantities. For use on 60-cycle half-wave rectification, f_c should be less than 60; for the full-wave rectifier f_c , it should be less than 120 cycles. The lower f_c , the better will be the filtering at the desired frequency as shown by the rising attenuation curve of Fig. 32a.

The load R is the usual starting place for finding the filter values when the voltage divider and tube load have been calculated previously. When the point on the R column is known, and f_c is, say, 50 cycles, the values of L and C are quickly determined. It is seen from Fig. 33 that, for a given cutoff frequency f_c , as the load resistance increases L also increases while C goes down. Very high load resistances require chokes of large inductance values, but as high-resistance loads mean small currents, the use of large inductances may be feasible.

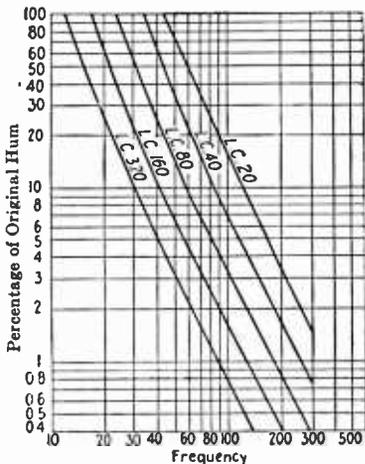


FIG. 34. Smoothing effected by various products of inductance (henrys) and capacity (microfarads).

secured concerning the amount of filtering needed in the circuit shown in Fig. 37. Suppose the output stage is supplied with plate power which is filtered x per cent (so that its hum is reduced to x per cent of its unfiltered value), and at this value it gives no objectionable hum in the loud-speaker. Suppose further that the amplification between the plate of this last tube and that of the preceding tube is A . Then the preceding stage must have its plate supply filtered x/A per cent. This means that the plate supply of the next to the last output stage can have as a maximum only $1/A$ as much ripple as the output stage, because of the amplification. Figure 34 gives this relation in useful graphic form. If the stage of amplification has a voltage gain of 25, it is essential that the preceding tube be supplied with plate power with a maximum of one twenty-fifth the ripple, or 4 per cent.

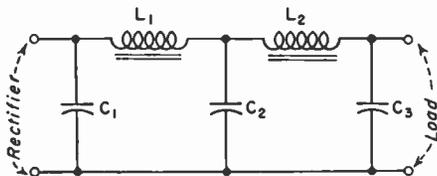


FIG. 35. Low-pass filter circuit.

Taking the frequency as 50 cycles, the nearest curve shows 3 per cent for an LC product of 320, which could be obtained with a 32 h choke and a 10- μ f capacitor.

32. RC Filter. A circuit similar to Fig. 37, using more resistors than chokes, is frequently used to provide an extra amount of filtering for stages preceding a power stage, as in Fig. 36. This is particularly useful when an output stage requires a high voltage and (1) when the voltage for other stages may or must be materially reduced, or (2) when the earlier stages are RC coupled, and low d-c power is required for them. The great advantage of a filter choke is its high impedance to the unwanted rectified (ripple) d.c., but a low resistance to the desired (filtered) d.c. If the amount of d-c

¹ Cocking, *Wireless World*, Nov. 19, 1930, pp. 565-568.

However, any low-pass filter using a choke is used mainly when currents range from an appreciable fraction of an ampere down to perhaps 10 ma. With currents lower than that it is often advisable to use an RC filter (see Sec. 32) since the power lost at low currents is negligible, and the cost of high-inductance chokes is proportionately too great.

Considering the filter for the amplifier of Fig. 9, to use the transformer designed in Sec. 29, the d-c load is 200 ma at 425 volts, $R = E/I = 425/0.2 = 2,125$ ohms. Taking f_c as 60 cycles, Fig. 33 shows $L = 10$ henrys, and $C = 3 \mu$ f. As noted previously, this is an approximate starting point. Common practice is to use a larger capacitor than is shown by this calculation. This practice will improve the filtering action, though not following the approximate design data of Eqs. (3) and (4).

31. Ripple per Stage. By assuming that the load resistance does not affect the values of L and C , a useful approximation¹ can be

potential is no great object, a resistance of as great a value as the impedance of the choke can be employed. If, as in Fig. 36, two stages of choke and capacitor filtering are used, the additional resistance and capacitance filter stages simply increase the amount of filtering without the chokes, which cost more, and are heavier than resistors. The RC values and the degree of filtering are given in Fig. 38 used in much the same manner as Fig. 34. The circuit of Fig. 36 is quite similar to that in Fig. 37 in eliminating the undesired feedback effects.

The chart of Fig. 33, based on Eqs. (3) and (4), gives satisfactory preliminary design data, but the experimental curves¹ showing the effects of load and different capacitors are quite interesting and give a clearer idea of the validity of the chart.²

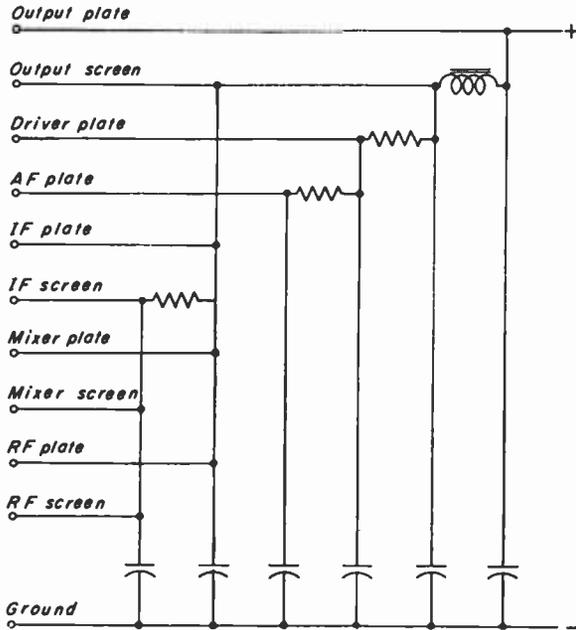


FIG. 36. Plate and screen supply which minimizes feedback.

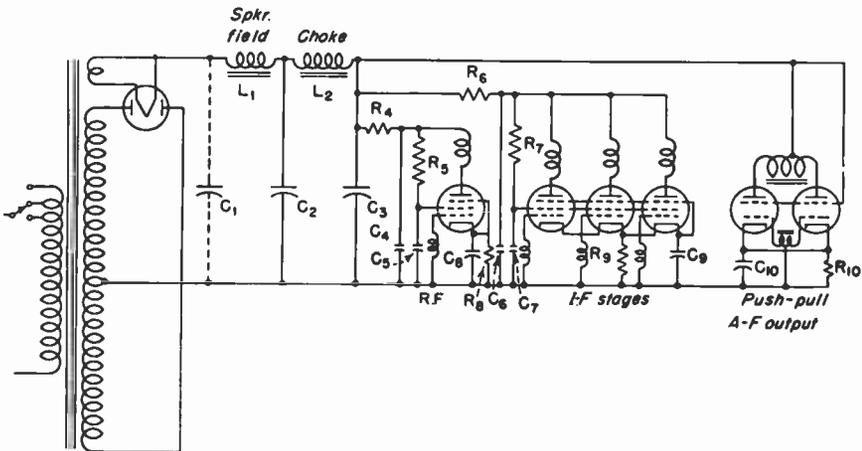


FIG. 37. Schematic receiver circuit showing graded filter.

33. First Filter Capacitor. The effect of the first filter capacitor, shown dotted in Fig. 37, is to raise the available output voltages. The addition or elimination of C_1

¹ Figures 39 through 42 are experimental curves taken from articles by Sidney Fishberg in *Aerovox Research Worker*.

² A theoretical calculation of the effects of C_1 , C_2 , and C_3 on the output voltage is given in *Gen. Elec. Rev.*, 19, 177, 1916.

gives, respectively, a capacitor-input, or a choke-input filter. Figure 39 gives the output voltage available as C_1 is varied from 0 to $4 \mu\text{f}$, as a function of load current. Figure 40 gives the per cent of ripple in output as C_1 is changed in value. This curve shows that the use of a single capacitor C_1 can never reduce the ripple much below

10 per cent with a reasonable value of capacitance. Much less than one-half of 1 per cent is needed in a good filter and, as at least two capacitors must be used to provide a filter section, this figure simply checks the theory.

34. Second and Third Filter Capacitors. Figure 41 gives the per cent ripple as a function of C_2 and C_3 for a given current (d-c load) drain. The most economical filter is obtained when $C_2 = C_3$. For example, suppose the ripple permissible to be 0.1 per cent. This can be furnished with $C_2 = 0$ if $C_3 = 5 \mu\text{f}$, a total of $5 \mu\text{f}$. But

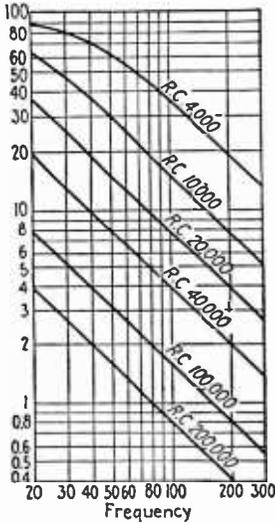


FIG. 38. Filtering effected by resistance (ohms)-capacitance (microfarads) circuit.

it can also be met with $C_2 = C_3 = 2 \mu\text{f}$, a total of $4 \mu\text{f}$. The dotted line gives the ripple value when $C_2 = C_3$. These figures apply only to the specific filter shown, but the relations between capacitor values hold for similar filter circuits. Note that these capacitances are rather low in total value. These data were obtained when paper capacitors were the rule, and higher capacitances in paper capacitors were then, as they are now, expensive. Considerable information applicable to the high-capacitance electrolytic capacitors can be gained from the figure; the basic theory is the same.

Figure 42 shows that the higher the values of C_2 and C_3 , the lower the percentage hum. It should be remembered that increasing current means a decreasing load resistance. From Fig. 33; assuming f_c to be constant, the capacitance should increase and the inductance decrease as the load resistance decreases. Thus, as Fig. 42 was taken using the same inductance coils throughout, larger values for C_2 and C_3 are needed as the current drain increases. It is almost certain that the inductance values of a given choke decrease as the current through it increases. To a certain extent this inductance decrease does not interfere with filtering, especially if the capacitance is increased as, referring again to Fig. 33, when the resistance decreases to half a given value, the capacitance should be doubled, while the inductance need only be half its initial value, if f_c is the same. Thus in Fig. 42, as in the other figures, the experimental facts agree with the theoretical chart, Fig. 33, and Eqs. (3) and (4), for this type of filter.

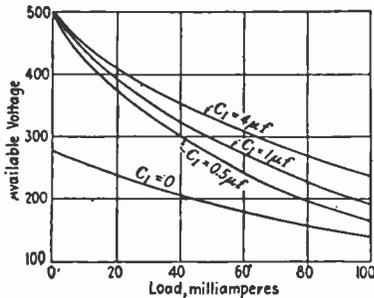


FIG. 39. Effect of C_1 on voltage available.

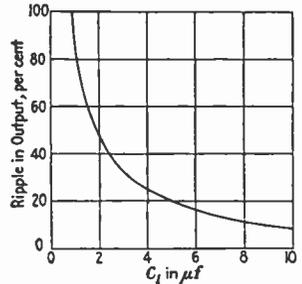


FIG. 40. Effect of C_1 on ripple in output.

35. Swinging Choke. A "swinging" choke is often used in transmitter circuits as the first choke following the rectifier. The inductance of this type of choke varies from a maximum greater than that shown to be needed at no load, to a minimum equal to that shown by calculation to be required for full load. This principle can be illustrated by a certain choke having 5-h inductance at 200-ma rated load. When the current decreases to 50 ma, the inductance value "swings" up to 10 henrys. The 5-h value is needed at full load—at no load the 10-h value is simply higher than needed.

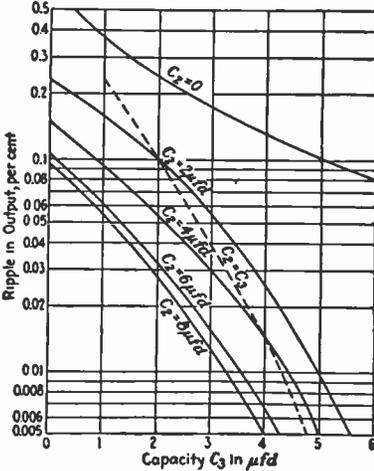


Fig. 41. Percentage ripple as a function of \$C_2\$ and \$C_3\$.

the current decreases to 50 ma, the inductance value "swings" up to 10 henrys. The 5-h value is needed at full load—at no load the 10-h value is simply higher than needed.

36. Filter Chokes Having Mutual Inductance. A scheme for decreasing filter weight is to have two consecutive chokes, e.g., \$L_1\$ and \$L_2\$ of Fig. 35, arranged to have a single core.

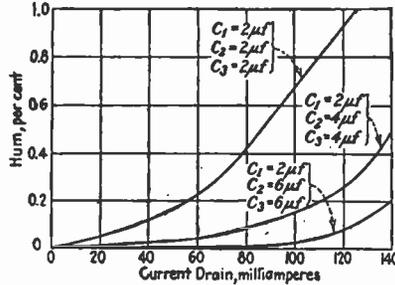


Fig. 42. Percentage hum as a function of current drain.

Power from the rectifier is fed to the junction between \$L_1\$ and \$L_2\$ and \$C_2\$ is eliminated. The value of \$C_1\$ is varied to give maximum filtering. Experimental curves of a-c ripple and values of the condenser, as a function of the relative turns of \$L_1\$ and \$L_2\$, will be found in the reference.¹ Such schemes are seldom used now since electrolytic capacitors of high values are inexpensive, and two chokes are seldom used.

37. Design of Filter Chokes. It is important that the filter choke be designed to carry the desired d.c. and at the same time to offer the necessary reactance to the a-c component. A direct method of design² has been derived using both the normal and incremental permeability curves for the core material. The derivation gives the two following working equations:

$$\frac{LI^2}{V} = \frac{B^2 \left(\frac{1}{\mu} + \frac{a}{l} \right)^2 \times 10^{-8}}{0.4 \left(\frac{1}{\mu_\Delta} + \frac{a}{l} \right)} \tag{5}$$

$$\frac{NI}{l} = \frac{B}{0.4\pi} \left(\frac{1}{\mu} + \frac{a}{l} \right) \tag{6}$$

where \$L\$ = henrys

\$I\$ = d-c amp

\$V\$ = core volume, cu in.

\$N\$ = turns

\$l\$ = magnetic path, cm

\$a\$ = air gap, cm

\$B\$ = steady flux density in iron and air gap, gauss

\$\mu\$ = normal permeability \$B/H\$

\$\mu_\Delta\$ = incremental permeability \$\Delta B/\Delta H\$ for a minor hysteresis loop

¹ Proc. IRE, January, 1930, p. 161.

² Hanna, C. R., Jour. AIEE, 46, 128, February, 1927.

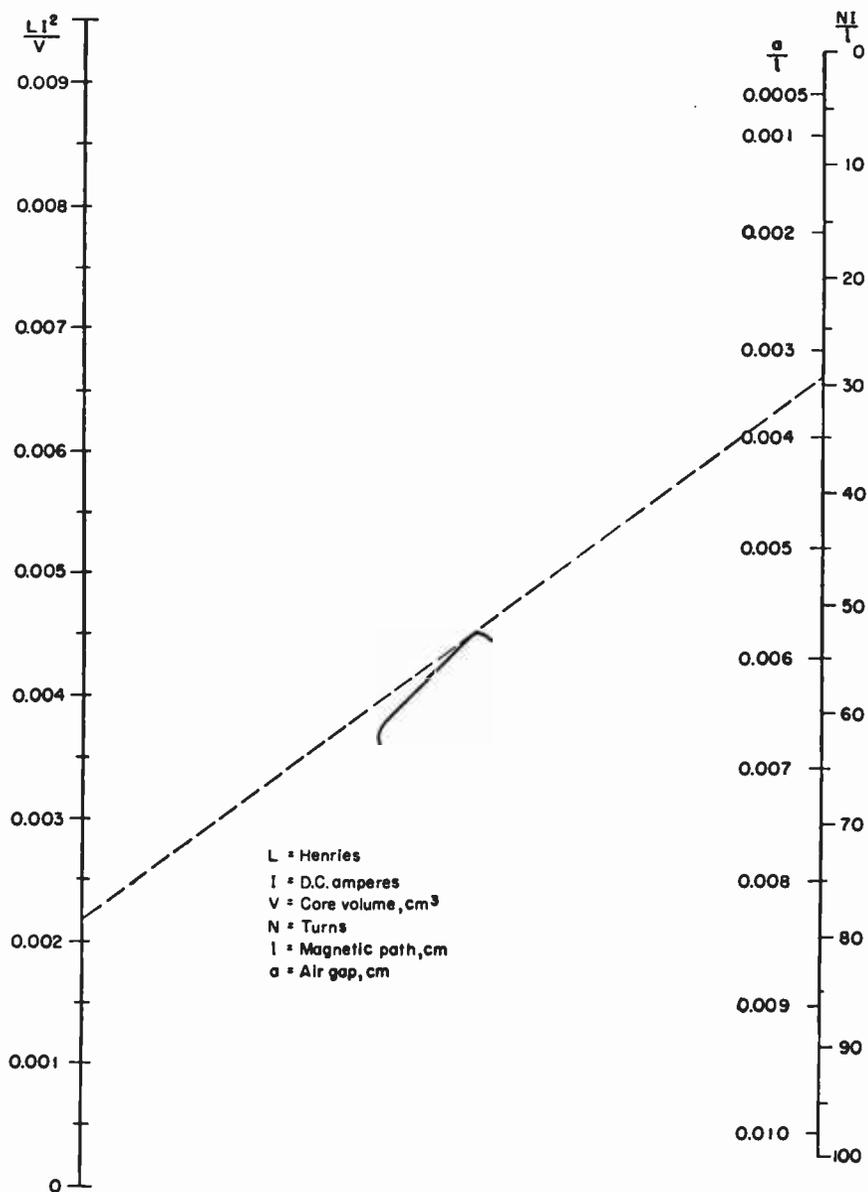


FIG. 43. Choke design: 4 per cent silicon core.

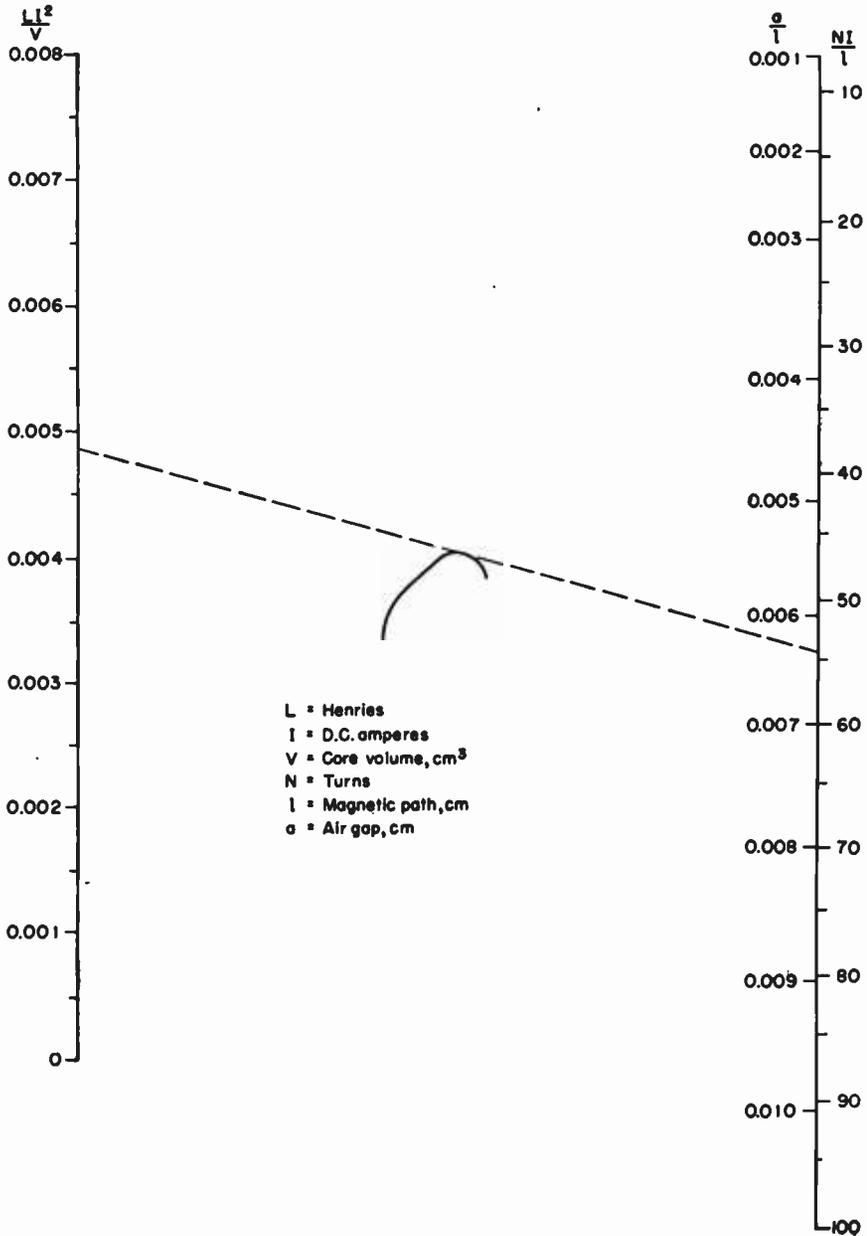


FIG. 44. Choke design: hipernik core.

The original curves were plotted with a/l as a parameter, LI^2/V being the ordinate, and NI/l as the abscissa for both 4 per cent silicon steel and hipernik. Figures 43 and 44 are alignment charts which include the data of the original curves.

Figure 45 gives typical permeability curves for five grades of magnetic material commercially available. A chart for calculating chokes, using Armco Radio 4 is in Fig. 46, the columns being similar to those of Figs. 43 and 44, and the procedure is the same.

38. Data for Calculating Filter Chokes. It has been suggested¹ for a choke design that the values be as follows:

	LI^2/V	NI/l	a/l
Initial choke.....	0.0031	38	0.0046
Following choke.....	0.0049	58	0.0062

These data agree closely with values of the charts of Fig. 43 for 4 per cent silicon iron and give a specific starting point for the calculation of a choke which carries d.c.

Figure 45 shows typical permeability curves for commercially available magnetic materials. Note that the materials with the highest μ at 6 kilogausses usually have lower μ at some value of B between 11 and 14 kilogausses than those with a lower maximum μ .

The ordinate of Fig. 45 shows permeability on the cgs system. Strictly this permeability only applies to B in kilogausses, and not to B in kilolines per square inch. The latter is added to the figure as a convenient reference.

39. Designing a Choke to Carry D.C. This choke is to carry 200 ma, L_1 of Fig. 35, to supply the amplifier of Fig. 9, using the 4 per cent silicon-iron core of Fig. 48. The resistance is to be 100 ohms, and it is

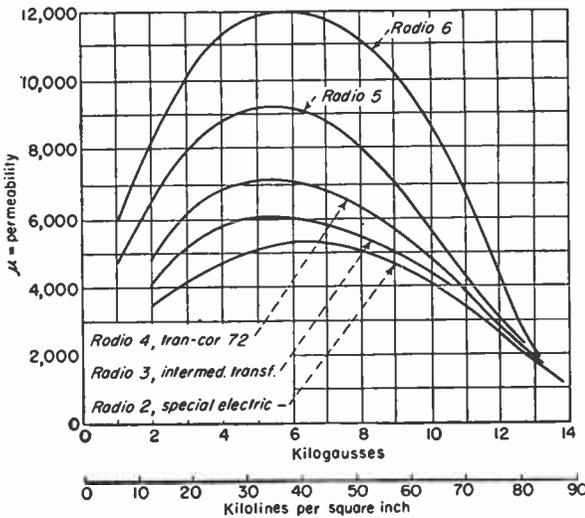


FIG. 45. Typical permeability curves of radio grades of Armco iron.

planned to use a round spool, Fig. 49, for its ease of coil winding. The data for use, including unknowns, are the following:

$$V = 46.5 \text{ sq cm} \times 2.29 \text{ cm} = 106.5 \text{ cc}$$

$$I = 0.200 \text{ amp}$$

$$l = 18.3 \text{ cm}$$

$$N = ? \text{ turns}$$

$$a = ? \text{ in., air gap } \left(\frac{a}{l} \text{ will be cm per cm} \right)$$

$$R = ? \text{ (preferably about 100 ohms)}$$

The following subparagraphs (1) and (2) show the use of Sec. 37 and Fig. 43, in finding L , N , and a ; and (3) shows the calculation of R .

¹ Smith, Langford, "Radiotron Designer's Handbook," 3d ed. p. 220, 1942, data revised to use same units as previous paragraphs.

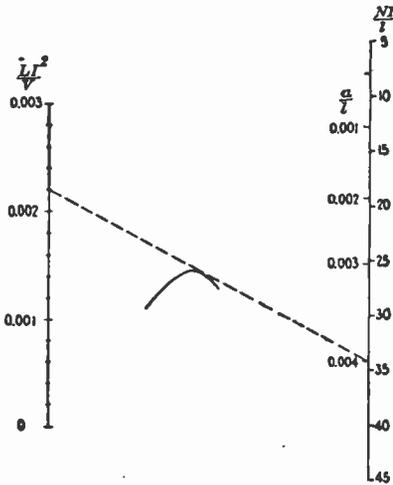


FIG. 46. Choke design: Armco Radio 4.

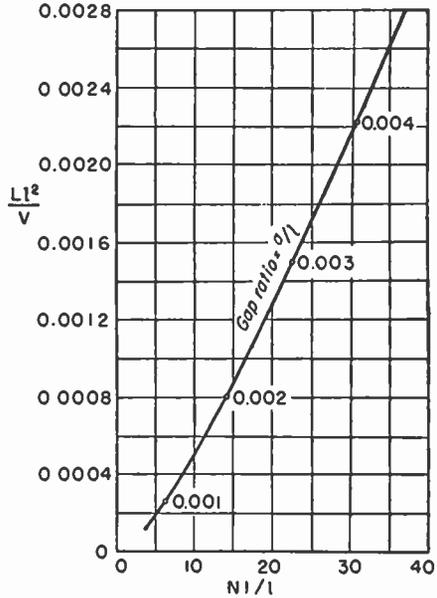
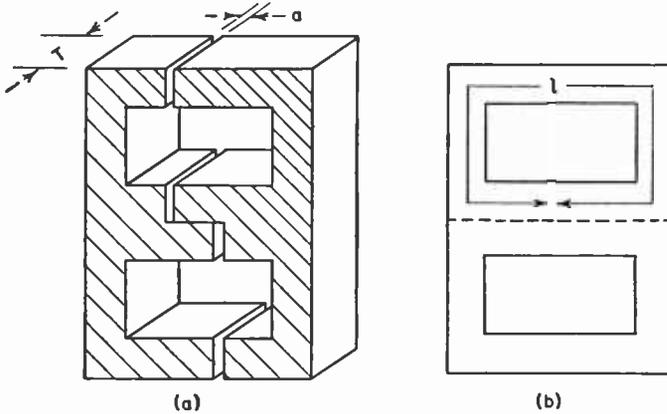


FIG. 47. Typical reactor design curve Armco Radio 6 iron. V in cc, l in cm.



Area of cross-hatched face = 7.2 sq in. = 46.5 sq cm
 Thickness of stack = $T = 0.9$ in. = 2.29 cm
 Length of magnetic path = $l = 7.2$ in. = 18.3 cm
 Air gap = a (cm or in.)
 Gap ratio = a/l (cm/cm or in./in.)

FIG. 48. Typical choke coil to carry d.c.

Note in Fig. 48 that the core has a lap construction in the center leg. The air gap is present in four places: top, bottom, and twice in the center leg. It is customary to neglect the two gaps in the center leg, since the upper and lower punchings touch mechanically; only the upper and lower gaps a are recognized, the a found by the calculation.

1. Using data from Sec. 37, the initial choke will use $LI^2/V = 0.0031$, $NI/l = 38$, and $a/l = 0.0046$. These are solved for L , N , and a :

$$L = \frac{0.0031V}{I^2} = 0.0031 \times \frac{106.5}{0.04} = 8.3 \text{ henrys}$$

$$N = \frac{38l}{I} = 38 \times \frac{18.3}{0.2} = 3,480 \text{ turns}$$

$$a = 0.0046l = 0.0046 \times 18.3 = 0.084 \text{ cm} = 0.033 \text{ in.}$$

2. An alternative is to use data from Fig. 43 for 4 per cent silicon steel. A calculation will be made for a 3,480-turn coil on the core of Fig. 48. The starting point, when turns and core are specified, is NI/l which is $3,480 \times 0.2/18.3 = 38$. This value from the right-hand line of Fig. 43 shows a/l to be 0.0046, thus

$$a = 0.0046 \times 18.3 = 0.084 \text{ cm} = 0.033 \text{ in.}$$

Lining a straightedge from $NI/l = 38$, tangent to the curve in the center of the figure,

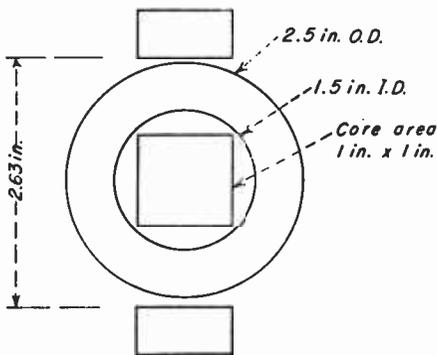


FIG. 49. Round spool for choke winding.

LI^2/V is found to be 0.0032; so $L = 0.0032 \times 105.5/0.04 = 8.4$ henrys. It should be mentioned that two of the figures, Figs. 44 and 47, portray data for high-quality magnetic core steel. Such materials are of considerable advantage in a-f chokes, when the frequency range is high (not shown) and when the induction is low, see Fig. 45 of permeability at 5 kilogausses, rather than for the 120-cycle full-wave filter choke of the example just shown.

3. The desired spool is Fig. 49, a section of the core of Fig. 48, clearance dimensions are adequate for assembling into the core. The requirement is 3,480 turns on this spool; outside diameter is 2.5 in., inside diameter 1.5 in., axial length 1.7 in. (not shown). The window area is the axial length times half the difference between outside and inside diameters; $1.70(2.5 - 1.5)/2 = 0.85$ sq in. Since the chart of Fig. 29 is based on turns per square inch, it is necessary to find the turns for a full square inch; $3,480 \text{ turns} \div 0.85 \text{ sq in.} = 4,100 \text{ turns per sq in.}$ Consulting Fig. 29, and recalling that a space factor of one-half is desirable, it is found that No. 30 regular enamel copper wire has a theoretical maximum of 8,500 turns per sq in., and 80 ohms per cu in. The space factor is $4,100/8,500 = 0.48$, or 48 per cent.

Before the resistance can be found, the winding volume is calculated:

$$V = \frac{\pi}{4} (D^2 - d^2) \times L = 0.7854(6.25 - 2.25)1.70 = 5.34 \text{ cu in.}$$

Figure 50 is a chart for calculating the area A , knowing D and d . The resulting area is to be multiplied by the axial length of the coil to get its volume. As 48 per cent of this space is filled with wire, the resistance of a No. 30 copper wire coil is $5.34 \text{ cu in.} \times 80 \text{ ohms per cu in.} \times 0.48 = 205 \text{ ohms}$. This result is more than twice the desired 100-ohm value, and a change is in order.

Since the number of turns is fixed, a larger wire must be employed, but whether or not to continue with the round spool of Fig. 49 is the question. It will be assumed that a reasonable space factor can be higher in a round coil, and Fig. 49 will still be used. Further calculation will show that a size between B & S No. 27 and No. 28

is needed to fit the conditions specified. Here an advantage of Fig. 29 is illustrated, data for fractional sizes being readily found.

Using B & S No. 27½, the turns per square inch = 4,800, and ohms per cubic inch = 23, space factor 4,100/4,800 will be 85 per cent, the resistance

$$5.34 \times 23 \times 0.85 = 104 \text{ ohms}$$

This is close to the required value.

However, a square-section coil spool, similar to that of Fig. 25 for the transformer, would allow 3,480 turns of No. 27 enamel wire to be put on with a resulting resistance of less than 100 ohms.

40. Impregnation of Transformer and Choke Coil Windings. The process and materials used in impregnation of coils of wire should¹ remove all moisture; fill up the pores and seal the winding against reentrance of moisture; improve the insulation

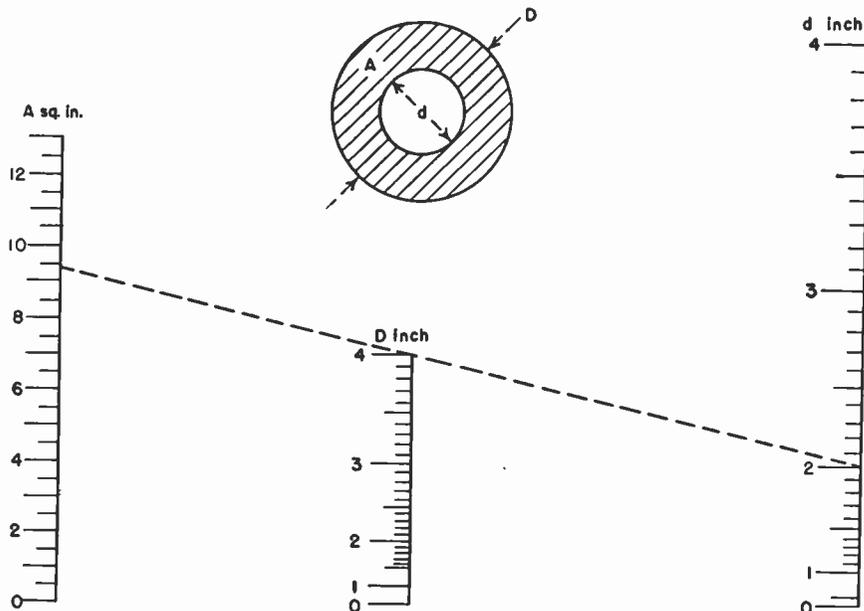


FIG. 50. Area of concentric winding spool. Multiply by coil length to get volume.

of the winding, especially cotton-covered wires; and help to hold the winding in place and prevent movement of the turns. The resins used for impregnation are composed of solids dissolved in volatile solvents. As much as 60 per cent resin can be dissolved in the solvent and still obtain a varnish of low enough viscosity to impregnate coils. The disadvantage is that voids are left inside the coil when the solvent has evaporated. A major advance² in electrical insulation is a varnish whose solvents react with the base resin, when heated, to make a final solid. Typical of these is Fosterite,³ a two-step process which makes a solid structure of the insulation, greatly improving the resistance to mechanical shock, and obtaining a cooler operating coil through more uniform and better heat conductivity. These advantages are available at a lower cost than for glass or other class B insulation. Coils treated with Fosterite can run

¹ Anaconda Wire and Cable Co., Handbook C-352, 4th ed., pp. 40-41, 1943.

² Foster, N. C., Styrene Copolymer Solvent Reactive Varnishes—"Fosterite," *Westinghouse Sci. Paper* 1243, preprint 90-24, The Electrochemical Society, presented October, 1946, Toronto, Canada.

³ Using materials M8689-1 and M8690-2 (Westinghouse Elec. Corp.).

continuously at 115 to 125°C. Such coils absorb negligible amounts of water, even after months of total immersion.¹

41. Filter Capacitor Ratings. Filament-type rectifiers come up to operating temperature in a few seconds and begin supplying rectified voltage before indirectly heated cathode tubes (load) are warm enough to take their rated currents. Dry-contact rectifiers also operate quickly, the warm-up time is zero. Thus peak voltages are used in specifying filter capacitor ratings. For example, the first capacitor should be rated to stand 1.41 times the rms transformer secondary voltage, for a 400-0-400 volt secondary the peak is $400 \times 1.41 = 465$ volts. (Half the secondary supplies peak voltage in a full-wave rectifier.) The second and third filter capacitor also may receive practically full peak voltage when no load current is drawn. For some circuits, an appreciable percentage of the total load current flows in a voltage divider, as a waste or "circulating" current. In this case, the second and third capacitor ratings can be proportionally lower by the amount of the voltage drop in the chokes or filter resistors. For example, in Fig. 9 there are two series resistors 1,400 and 5,000 ohms from +425 volts to ground. Until the 6L6 tubes warm up, no appreciable screen-grid current flows, but if a filament rectifier is furnishing this 425 volts, a current flows through the two resistors, $I = E/(R_1 + R_2) = 425/(1,400 + 5,000) = 0.079$ amp. The voltage applied to the screens and to the filter capacitor across the screens will receive $425 - IR_1 = 425 - 0.079 \times 1,400 = 425 - 110 = 325$ volts. Thus this capacitor maximum rating can be 325 volts, but as the screen takes current, the operating screen-grid voltage is lower, nearer 300 volts.

While it is true that good filter capacitors will stand voltages greater than their d-c ratings, the regular practice of applying such higher voltages is seldom advisable from a life standpoint. Electrolytics, especially, are subject to rapid deterioration if worked above their ratings.

42. Types of Filter Capacitors. The two general types of filter capacitors are paper and electrolytic. In general, the paper capacitor has no polarity and extremely low leakage current, changes little with wide temperature variations, is more expensive, weighs more, takes up more space for a given capacitance, and has long life. Paper capacitors are made for practically any high voltage, but seldom rated lower than 400 volts d.c. working.

Electrolytic capacitors are polarized, have appreciable leakage currents, somewhat restricted temperature ranges; single units are available for voltages up to 600 d.c. working, are relatively inexpensive, light in weight, and small in volume for a given capacitance. They have shorter life than paper capacitors. Reasonably sized electrolytics can be made with several thousand microfarads at such low working voltages as 15, 30, and 50.

Considerable progress in capacitor design and manufacture has been and is being made. Manufacturers should be consulted as to ratings and recommendations for capacitors for abnormal conditions of operation, such as extreme altitudes, high humidity, or unusual temperatures.

43. Electrolytic Filter Capacitors. The life and stability of an electrolytic capacitor are largely determined by the operating temperature of its electrolyte. The temperature depends on how much heat is generated inside the capacitor and the rate at which heat is taken away by the surroundings.² Electrolytic capacitors are normally rated for an ambient temperature of 60°C (140°F). Specially processed units can operate at 85°C (185°F). The low-temperature operating limit is seldom a problem for indoor use, but capacitors for use at temperatures as low as -40°C (-40°F) can be made with the effective rated capacitance as 40 per cent that at normal room temperature.

44. Low-temperature Operation of Electrolytic Capacitors.³ Changes in temperature cause less change in the capacitance of plain foil electrolytics at varying low tem-

¹ See also Reuben Lee, Recent Transformer Developments, *Proc. IRE*, **32**, 240, 1945.

² Sprague Electric Co., Catalogue 10A, pp. 23ff.

³ Aerovox Corp., catalogue, p. 123.

peratures than those using etched foil. However if small volume is important and changes of capacitance are not too serious, etched foil capacitors are desirable.

Electrolytic capacitors of 300 volts d.c. working have less change of capacitance than those with higher voltage ratings. Highly absorbent paper separators are used at 350 volts and lower, providing low series resistance and maximum electrolyte contact at the foils. For higher voltages a less absorbent separator is needed to withstand the voltage. High voltage ratings at very low temperatures can be secured by using several 300-volt electrolytics in series, each being shunted by a resistor (taking several times the leakage current) to ensure even division of the applied emf.

It may be repeated that only vacuum-type rectifier tubes are suitable for low-temperature operation. Mercury-type rectifiers operate in the range 24 to 60°C (75 to 140°F). See Secs. 11 and 15.

Typical specifications for electrolytic capacitors include rated capacitance and tolerance, operating frequency, rated voltage, normal operating voltage, peak operating voltage, operating temperature range (max and min), type of circuit used (full-wave, half-wave, doubler, etc.).¹

45. Paper Filter Capacitors. Paper filter capacitors are usually oil-filled or wax-impregnated. The selection of the proper oil or wax depends largely on the temperature range to be encountered, either in use or idle. Complete specification for a paper-dielectric capacitor requires: capacitance and tolerance; maximum instantaneous peak voltage; a-c ripple frequency; half-wave or full-wave; average ambient temperature; maximum and minimum temperature expected when not operating; special operating conditions such as high humidity, dusty location, high elevation.

Generally, a maximum operating temperature of 65°C is assumed although the voltage ratings are based on 20°C, standard atmospheric pressure (29.92 in. Hg), and 50 per cent relative humidity. Standard tolerances for capacitors of 0.1 μ f and over are -10 to +20 per cent.

¹ Sprague Electric Co., Catalogue 10A, p. 26.

CHAPTER 8

AUDIO-FREQUENCY AMPLIFIERS

BY GLENN KOEHLER¹

CLASSIFICATION AND INTRODUCTORY THEORY

1. Classification. An a-f amplifier is usually defined as one which is to work in the range of frequencies from 20 to 20,000 cps. Amplifiers for this purpose may be either selective or nonselective; *i.e.*, they may be made to amplify substantially a single frequency or a range of frequencies. Ordinarily the terminology implies that the amplifier will work over a range of frequencies.

There are two main classifications for a-f amplifiers. The first refers to the manner in which the tubes are operated with respect to their I_b - E_b characteristics. Under this classification a-f amplifiers are divided into class A, class AB, and class B.

A class A amplifier operates in such a manner with respect to the I_b - E_b characteristics that the output wave form for a single tube and any kind of an output impedance is substantially the same as the input wave form. Hence the operation is such that plate current flows for the full 360 electrical degrees of the input emf. When the grid of a class A amplifier is not driven positive with respect to the cathode, it is designated as class A₁. When the grid is driven positive, it is designated as class A₂. Class A₂ is seldom used because of high plate losses. Hence class A usually means that the grid is not driven positive.

A class AB amplifier is operated with sufficient grid bias for a given d-c plate potential that plate current flows for more than 180 electrical degrees but less than 360 deg of the input emf. This mode of operation requires two tubes in push-pull in order that the output wave form be nearly like the input wave form. When the grids of the tubes are not driven positive with respect to the cathodes, the designation is class AB₁. When the grids are driven positive, the designation is class AB₂.

A class B amplifier is operated with sufficient grid bias to reduce the plate current almost to zero when no input voltage is applied. Plate current flows for only 180 deg of the input cycle. It requires two tubes in push-pull to produce an output wave form that is nearly like the input wave form.

The second classification of a-f amplifiers refers to the output requirements.

An amplifier is called a *voltage amplifier* when the primary purpose is to obtain as much output voltage as possible with little regard to the efficiency of power conversion in the plate circuit.

An amplifier is called a *current amplifier* when the primary purpose is to obtain as much output current as possible with little regard to the efficiency of power conversion in the plate circuit.

An amplifier is called a *power amplifier* when the primary purpose is to obtain as much output power as possible with a limited amount of distortion and a reasonable efficiency of power conversion in the plate circuit.

There are many subclassifications under the second classification. For example, under voltage amplifiers there are single-ended amplifiers, push-pull amplifiers, single-stage amplifiers, multistage amplifiers, feedback amplifiers, phase inverter amplifiers, direct-coupled amplifiers, resistance-capacitance coupled amplifiers, transformer-coupled amplifiers, etc. Most of these subclassifications are applicable also to current and power amplifiers.

¹ Department of Electrical Engineering, University of Wisconsin.

2. General Requirements. An a-f amplifier must satisfy the following general requirements:

1. The gain of the amplifier must conform to a certain amplification-frequency characteristic.
2. The output wave form must not contain more than a certain amount of distortion that is generated in the amplifier itself.
3. The gain of the amplifier must be such that a certain output power is obtained from a given input voltage.
4. The noise and hum level of the amplifier should be within a preassigned limit.
5. The gain should not vary much with the usual variations in d-c operating voltage, temperature of filaments, etc.
6. The input and output conditions should be such as to work the amplifier out of a certain source impedance into a certain load impedance.
7. For certain applications, *e.g.*, where the observation of wave shapes is involved, the phase distortion should not exceed a certain amount.

3. Elements of an A-f Amplifier. The a-f amplifier tube acts as a power converter taking continuous power from the battery or d-c source in the plate circuit and converting this power into a-c power. The converted power is used to set up a voltage across an impedance in the plate circuit for the case of a voltage amplifier, or to supply power to a load for the case of a power amplifier. For carrying out this function, each stage of an amplifier must be furnished with an input coupling device, an output coupling device, and the necessary sources of power to actuate the tube. For the case of a multistage amplifier the input coupling device of one tube may be the output coupling device of the tube ahead of it.

4. Voltage Amplification per Stage. 1. *Simple Theory.* A single triode amplifier is shown in Fig. 1. The voltage-amplification theory given below applies to a tube of three elements or more when operated as a class A amplifier without external impedances in any of the elements other than the anode or plate circuit. In the simple theory the interelectrode capacitances of the tube and socket are neglected.

The two important constants of the amplifier tube are the amplification constant μ and the plate resistance r_p . The tube acts as a source of alternating emf which is controlled by the input voltage e_g . This equivalent source which has a voltage $-\mu e_g$ and an internal impedance r_p sets up a.c. in the external impedance Z_o . The a.c. through Z_o produces an alternating voltage across Z_o which is the output voltage e_o . The voltage amplification, or voltage gain of the amplifier is

$$A = \frac{E_o}{E_g} = \frac{-\mu Z_o}{r_p + Z_o} \tag{1}$$

In this expression $Z_o = R_o + jX_o$ and E_o and E_g are the vector values of e_o and e_g . Voltage amplification is also a vector quantity. The voltage E_g is used as the reference vector. Figure 2 shows the voltage amplification in per cent of μ plotted against ratios of output impedance to plate resistance: for cases where Z_o is a resistance R_o , or a reactance X_o .

Because of the approximate way in which the ear responds to sound, *i.e.*, logarithmically, it is convenient to express the gain of an amplifier logarithmically. The unit is the decibel, which is equal to 20 times the common logarithm of the absolute value of the voltage ratio. Hence the gain in decibels is $20 \log_{10} |A|$. The power gain in decibels can be determined from the voltage in decibels, only when the input

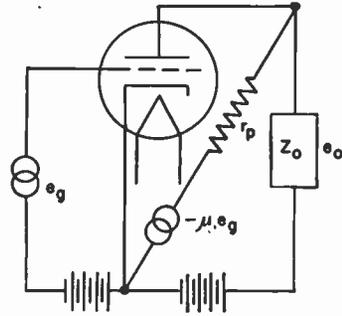


FIG. 1.—Triode amplifier.

and output impedances are known. Strictly speaking, the power gain in decibels is the more fundamental quantity.

Equation (1) is valid for pentode tubes as well as triode tubes when there are no impedances in any of the electrode circuits except the plate circuit. However, for a pentode used as a voltage amplifier, it is not practical to have the output impedance

Z_o comparable in magnitude to the plate resistance. Hence \dot{A} , in a form more suitable for a pentode voltage amplifier, is

$$\dot{A} = \frac{-g_{p0}\dot{Z}_o}{1 + \dot{Z}_o/r_p} \cong -g_{p0}\dot{Z}_o \quad (2)$$

because $\dot{Z}_o/r_p \ll 1$.

2. *Effects of the Interelectrode Capacitance.* The location of the interelectrode capacitances for a triode is shown in Fig. 3. These capacitances should include the tube itself and the socket. The capacitances given in the tube handbooks and manuals are usually for the tube alone. In many cases the socket interelectrode capacitances are as large as for the tube alone. When

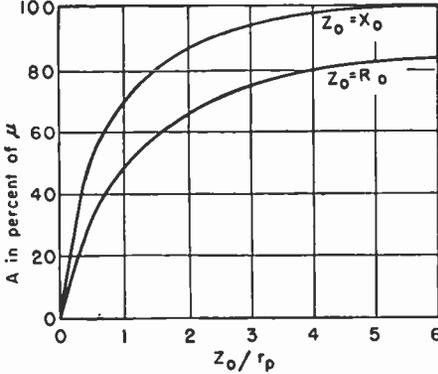


FIG. 2. Voltage amplification of a triode.

the socket capacitances are not known, it is good practice to add about $4 \mu\text{mf}$ for adjacent electrodes and $3 \mu\text{mf}$ for all others except in the case where the grid comes out the top which requires no change from that given in the handbook. Multigrid tubes used as class A triode amplifiers are treated similar to the triode when there are no impedances in any of the other grid circuits.

The voltage amplification \dot{A} for the circuit of Fig. 3 is

$$\dot{A} = \frac{\dot{E}_o}{\dot{E}_i} = \frac{j\omega C_{op} - g_{p0}}{g_p + \dot{Y}_o + j\omega(C_{op} + C_{pk})} \quad (3)$$

in which $\dot{Y}_o = 1/Z_o$, $g_{p0} = \mu/r_p$, and $g_p = 1/r_p$.

For a pentode voltage amplifier capacitance C_{pk} in Eq. (3) must be replaced by the output capacitance C_o which is equal to the plate-to-cathode capacitance plus the plate-to-suppressor capacitance plus the plate-to-screen grid capacitance.

Usually the susceptance ωC_{op} , for both pentodes and triodes, is very much smaller than g_{p0} in the a-f range and may, therefore, be neglected in the numerator of Eq. (3).

5. *Input Impedance.* The input impedance of the tube shown in Fig. 3 is the voltage \dot{E}_i divided by the current \dot{I}_i that would flow in the external grid circuit. For a high-vacuum tube, when operated so that the grid never goes positive, the current \dot{I}_i would be the vector sum of the currents through the capacitances C_{op} and C_{pk} . Since these two branches are effectively in parallel, it is better to consider input admittances. The expression for the input admittance is

$$\dot{Y}_i = j\omega C_{pk} + j\omega C_{op}(1 - \dot{A}) \quad (4)$$

The impedance \dot{Z}_i is the reciprocal of \dot{Y}_i . The voltage amplification \dot{A} is a vector quantity and is obtained from Eq. (3) or (1) when the interelectrode capacitances are negligible in their effects upon \dot{A} . When the output impedance is a resistance, the

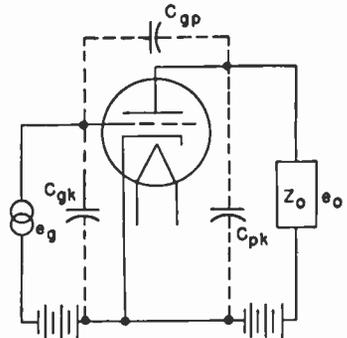


FIG. 3. Triode amplifier showing interelectrode capacitances.

value of A is usually a negative real quantity, and the capacitance C_{op} is multiplied by $(1 + |A|)$. Under certain conditions when the impedance Z_o has an inductive reactance, the input impedance Z_i is made up of a capacitive reactance and negative resistance. This is an important consideration in an a-f amplifier because it may cause sustained oscillations which in turn may cause very bad distortion.

For pentode voltage amplifiers the capacitance C_{pk} in Eq. (4) is replaced by the input capacitance C_i which includes the capacitances between the grid and all the other electrodes except the plate. The quantity $\omega C_{op}(1 - A)$ is usually so small compared to ωC_i for a pentode that it may be neglected.

The input impedance of an amplifier tube is an important consideration when designing multistage amplifiers. As a general rule this impedance plays a part in the performance of a voltage amplifier for all frequencies above about 3,000 cps.

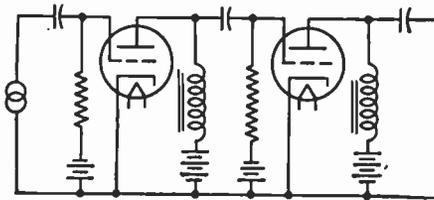


FIG. 5. Impedance-capacitance-coupled amplifier.

1. Resistance-capacitance coupled amplifier, illustrated in Fig. 4.
2. Impedance-capacitance coupled amplifier, illustrated in Fig. 5.
3. Transformer-coupled amplifier, illustrated in Fig. 6.

There are several variations of the second class. The resistances in the grid circuits may be replaced by inductive impedances. In general the elements in both the plate and grid circuits may be any type of impedance as long as they pass d.c. The more common types are the one shown and the one with simple inductive impedances in both the plate and the grid circuits. A single multistage amplifier may be a combination of these different fundamental types.

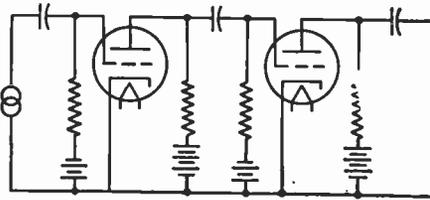


FIG. 4. Resistance-capacitance-coupled amplifier.

THEORY AND DESIGN OF CLASS A MULTISTAGE VOLTAGE AMPLIFIERS

6. Methods of Coupling. Multistage class A voltage amplifiers are usually divided into three classes:

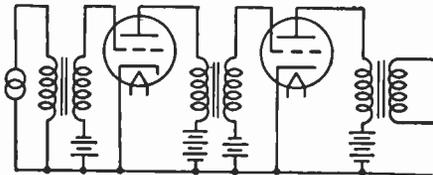


FIG. 6. Transformer-coupled amplifier.

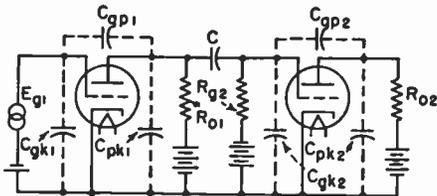


FIG. 7. Resistance-capacitance amplifier showing tube capacitances.

At the low frequencies the coupling capacitor C causes the amplification to decrease with decrease in frequency because there is a voltage drop, in C , from the plate of tube 1 to the grid of

7. RC-coupled Amplifier. This class of multistage amplifiers is illustrated in Fig. 7, with the interelectrode capacitances of the tubes shown in dotted line. Consider the voltage amplification of stage 1, i.e., E_{p2}/E_{g1} . Over a middle range of frequencies the voltage amplification is substantially independent of the frequency; neither the coupling capacitor nor the interelectrode capacitances have any effect. At the low

tube 2 which increases with decrease in frequency. At the high frequencies the interelectrode capacitances cause the amplification to decrease with increase in frequency because these capacitances lower the impedance in the external plate circuit of tube 1.

8. Frequency Characteristic. The medium-frequency gain A_M of stage 1 is

$$\frac{E_{o2}}{E_{o1}} = A_M = \frac{g_{op1}}{G_{o2} + G_o + g_{p1}} \tag{5}$$

in which

$$g_{p1} = \frac{\mu_1}{r_{p1}} \quad G_{o2} = \frac{1}{R_{o2}} \quad G_o = \frac{1}{R_{o1}} \quad \text{and} \quad g_{p1} = \frac{1}{r_{p1}}$$

$20 \log_{10} A_M$ will be used as the reference level, or zero level, to show what happens at low and high frequencies. The l-f gain, A_L , in terms of medium-frequency gain, is

$$A_L = \frac{A_M}{\sqrt{1 + (G_e/\omega C)^2}} \tag{6}$$

in which $G_e = \frac{G_{o2}(G_o + g_{p1})}{G_o + g_{p1} + G_{o2}}$ and C is the coupling capacitance between stages 1 and 2.

The loss at low frequencies, due to C , is equal to $20 \log_{10} \sqrt{1 + (G_e/\omega C)^2}$. The curves of Fig. 8 show the relation between C and G_e for particular decibel losses at a frequency of 50 cps. The curves may be used to predict the decibel loss due to C at

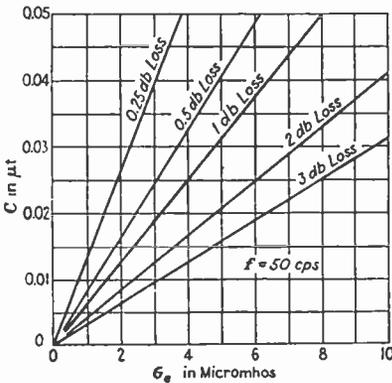


FIG. 8. Loss in low-frequency amplification due to coupling capacitor.

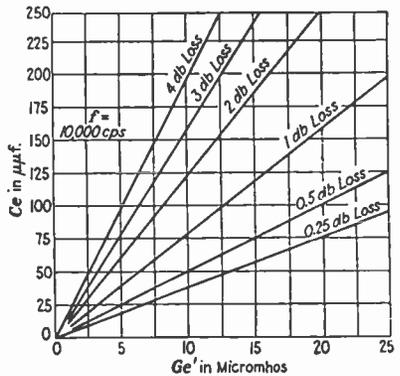


FIG. 9. Loss in high-frequency amplification due to interelectrode capacitances.

any other frequency f_z by multiplying the ordinates by $50/f_z$ and locating the known value of C on the new scale. Both scales may be changed simultaneously by multiplying by a factor x to provide a more convenient range for G_e . To illustrate the use of the curves, suppose $r_{p1} = 100,000$, $R_{o1} = 200,000$, and $R_{o2} = 500,000$, then $G_e = 1.76 \times 10^{-6}$. For 0.5 db loss at 50 cps, a coupling capacitor C equal to $0.0125 \mu f$ is required.

The h-f gain, A_H , is

$$A_H = \frac{A_M}{1 + (\omega C_e/G_e')^2} \tag{7}$$

in which for triodes $C_e \cong C_{op1} + C_{pk1} + C_{ok2} + C_{op2}(1 + |A_{v2}|)$ (see Fig. 7), and $G_e' = G_{p1} + G_o + G_{o2}$.

For a pentode $C_e = C_{o1} + C_{i2}$ where C_{o1} and C_{i2} are the output and input capacitances of tubes 1 and 2, respectively.

The loss due to the shunting action of the effective capacitance C_e at the high fre-

quencies is $20 \log_{10} \sqrt{1 + (\omega C_e/G_e')^2}$. The curves of Fig. 9 show the relation between C_e and G_e' for various decibel losses at a frequency of 10,000 cps. For a frequency f_x , multiply the present ordinates by $10,000/f_x$ and locate the capacitance C_e on the new scale. Suppose C_e is equal to $84 \mu\mu f$, then for the values given in the example above $G_e' = 10^{-6}$ and the loss at 10,000 cps is about 0.5 db.

Figure 10 illustrates the gain vs. frequency characteristics of a typical RC-coupled amplifier. In this figure two frequencies are designated f_L and f_H . These are the frequencies at which the gain has dropped to 0.707 times the mid-frequency gain. They are sometimes called the half-power frequencies because the power output of the output tube at these two frequencies is only one-half of the power at the mid-frequency. At these two frequencies the voltage gain of the stage is 3 db below mid-frequency gain. For the frequency $f_L G_e/\omega C = 1$ and for $f_H \omega C_e/G_e' = 1$. The phase shift is 45 deg lead at f_L and 45 deg lag at f_H .

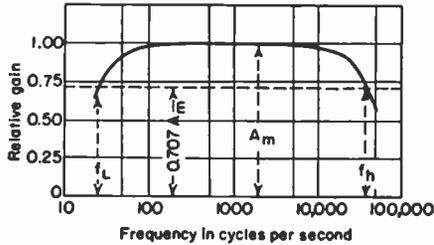


FIG. 10. Typical gain-frequency characteristics of resistance-capacitance-coupled amplifier.

In an amplifier of this type there is some phase distortion at both the highest and the lowest frequencies which the amplifier will pass without appreciable loss. The change in the phase angle of the voltage amplification with the frequency is at low frequencies $\theta_L = \tan^{-1} G_e/\omega C$, and at high frequencies it is $\theta_H = \tan^{-1} \omega C_e/G_e'$.

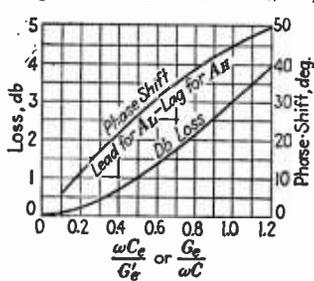


FIG. 11. Decibel loss and phase shift for resistance-capacitance-coupled amplifier.

The phase shift in the amplifier, i.e., the angle of departure of A_v from 180 deg, is illustrated by Fig. 11. This figure also shows how the decibel loss below the gain at the medium frequencies depends upon $\omega C_e/G_e'$ at the high frequencies and $G_e/\omega C$ at the low frequencies. For the frequency given by

$$f = \frac{1}{2\pi} \sqrt{\frac{G_e G_e'}{C C_e}} \tag{8}$$

the phase shift of the amplifier is zero. This may be considered the mid-frequency. The gain at this frequency is substantially the same as that given by Eq. (6).

9. Design of a RC-coupled Amplifier. When considering the proposition of using a certain tube in stage 1 (Fig. 7), to drive tube 2 and also give a preassigned amount of gain for stage 1, the first question is what will be the response at the highest frequency to be amplified. This is settled by determining the effective capacitance C_e (it is assumed that A_2 is known) and using the curves of Fig. 9 to find the value of G_e' for the allowable loss at the highest frequency. This value of G_e' will determine the medium-frequency gain of stage 1 [see Eq. (6)]. In calculating C_e the interelectrode capacitances given in the tube handbooks and manuals must be increased by 3 to $5 \mu\mu f$ to include the socket and other stray capacitances except for the electrode that comes out the top of the tube.

To determine the size of C for a preassigned response at the lowest frequency, it is necessary at this point in the design to fix the size of G_{g2} , or R_{g2} , and G_o , or R_o . The following considerations are pertinent to fixing the sizes of these resistors. It is always well to use as small a coupling capacitor as possible. Hence, because of the way that C depends upon G_{g2} , R_{g2} should be as large as possible but should not exceed the maximum recommended value given in the tube tables. In any event the product of $R_{g2}C$ should not exceed approximately 0.05 because of the tendency of C to become charged from a very small grid current and thereby cause the grid bias to become shifted.

For a given value of $G_o + G_{o2}$, and this is fixed when G_o' is fixed for a given tube, it is well to make R_o somewhat higher than the plate resistance r_{p1} to reduce distortion if the tube is worked very hard. On the other hand, R_o lowers the d-c voltage which must be supplied by the plate-voltage source.

All of the above theory and design is valid for pentode tubes as well as triode tubes. Pentode tubes are generally superior to high- μ triode tubes for RC amplifiers. The effective capacitance C_e is smaller and the gain per stage is generally higher for pentodes. Consequently a higher gain or a greater frequency range can be obtained by the use of pentode tubes. However a triode gives less distortion when the output voltage is large. A triode should generally be used to drive a class A power amplifier.

Grid bias for each of the tubes in RC amplifiers is usually obtained from a resistor, shunted by a capacitor, in the cathode circuit instead of batteries as shown in Fig. 7. The effects of cathode current bias are discussed in Sec. 16. Direct voltage for the screen of a pentode is generally obtained by connecting a resistor between the screen and the d-c supply for the plate. A by-pass capacitor is then connected between the screen and the cathode. The effect of this method for obtaining screen voltage is discussed in Sec. 16.

10. Impedance-Capacitance-coupled Amplifier. Under this classification of multi-stage amplifiers would fall almost any type of coupling except transformer coupling.

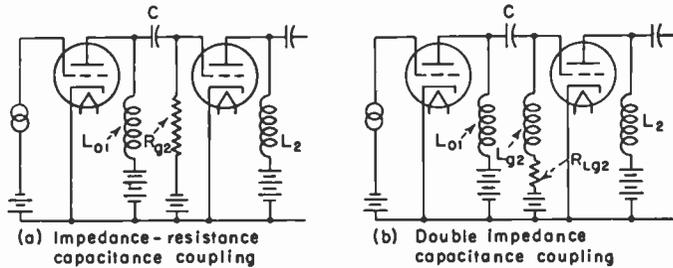


FIG. 12. Impedance-capacitance-coupled amplifiers.

RC coupling has special characteristics and is therefore treated in Sec. 8. The usual accepted types of the classification herein discussed are the two shown in Figs. 12a and b. The type shown in Fig. 12b is sometimes called double-impedance coupled. These types have frequency characteristics inferior to the RC amplifier but possess some other advantages. For example, it requires less B supply voltage to give the same plate voltage because of the much lower d-c voltage drop in the plate circuit. By a double-impedance scheme the gain at the low frequencies can be made higher than the gain at intermediate frequencies. This is sometimes useful in frequency-response equalization.

For the type shown in Fig. 12a the voltage amplification for stage 1 at medium frequencies is

$$A_M = \frac{E_{o2}}{E_{o1}} = \frac{g_{p01}}{G_{p1} + G_{o2}} = \frac{\mu_1 R_{o2}}{r_{p1} + R_{o2}} \tag{9}$$

in which $g_{p01} = \mu_1/r_{p1}$ and $g_{p1} = 1/r_{p1}$ for the tube of stage 1 and $G_{o2} = 1/R_{o2}$. In some cases it may be necessary to add the core-loss conductance for L_{o1} to G_{o2} . The voltage amplification at low frequencies in terms of A_M is rather involved. It is

$$A_L = \frac{A_M}{\sqrt{1 + \left(\frac{r_{p1} R_{o2}}{r_{p1} + R_{o2}}\right)^2 \left[\frac{1}{\omega^2 L_{o1}^2} \left(1 + \frac{1}{R_{o2}^2 \omega^2 C^2}\right) - \frac{2}{R_{o2}^2 \omega^2 L_{o1} C} + \frac{1}{r_{p1}^2 R_{o2}^2 \omega^2 C^2} \right]}} \tag{10}$$

When $C \cong 0.05 \mu f$ and $R_{g2} \cong 0.5$ megohm and $f \cong 50$ cps, this equation reduces to

$$A_L \cong \frac{A_M}{\sqrt{1 + \frac{1}{\omega^2 L_{o1}^2} \left(\frac{r_{p1} R_{g2}}{r_{p1} + R_{g2}} \right)^2}} \tag{11}$$

From Eq. (11) it is seen that there is a loss in amplification at the low frequencies. The loss in amplification in decibels due to insufficient reactance in choke L_{o1} is equal to

$$20 \log_{10} \sqrt{1 + \frac{1}{\omega^2 L_{o1}^2} \left(\frac{r_{p1} R_{g2}}{r_{p1} + R_{g2}} \right)^2}$$

At the high frequencies the voltage amplification, A_H , is

$$A_H = \frac{A_M}{\sqrt{1 + \left(\frac{\omega C_e}{G_e} \right)^2}} \tag{12}$$

in which C_e is the effective capacitance due to the tubes (see Sec. 8), plus the distributed capacitance of the choke, and G_e equals $G_{p1} + G_{g2}$ plus a conductance $1/R_c$ due to the core loss of the choke. The relation between C_e and G_e at 10,000 cps is the same as that given by the curves of Fig. 9. (See explanation in Sec. 8 for extending the range of the curves.)

The type of amplifier illustrated in Fig. 12b has some interesting characteristics. The medium-frequency amplification is $A_M = \mu_1$, neglecting the core losses of the two coils. For the case in which ωL_{o1} is several times R_{Lg2} and is at least three times r_{p1} , the amplification per stage at low frequencies in terms of that at medium frequency is

$$A_L = \frac{A_M}{\sqrt{\left(\frac{f_r}{f} \frac{1}{Q} \right)^2 + \left(1 - \frac{f_r^2}{f^2} \right)^2}} \tag{13}$$

where $f_r = \frac{1}{2\pi \sqrt{L_{g2} C}}$

$$Q = \frac{\omega_r L_{g2}}{r_{p1} + R_{Lg2}}$$

Using the medium-frequency gain as the reference and plotting

$$20 \log_{10} \sqrt{\left(\frac{f_r}{f} \frac{1}{Q} \right)^2 + \left(1 - \frac{f_r^2}{f^2} \right)^2}$$

as ordinates and f/f_r as abscissas for various values of Q , the curves of Fig. 13 result. These curves explain the characteristics of this type of coupling and furnish quantitative information on how to fix the values of L_{g2} and C for a particular performance at the low frequencies. At the frequency f , the gain, or loss, in decibels is equal to $20 \log_{10} Q$. The curves also show how the gain, or loss, varies with the frequency f for a particular case. The phase distortion at low frequencies would be very bad for an amplifier of this type.

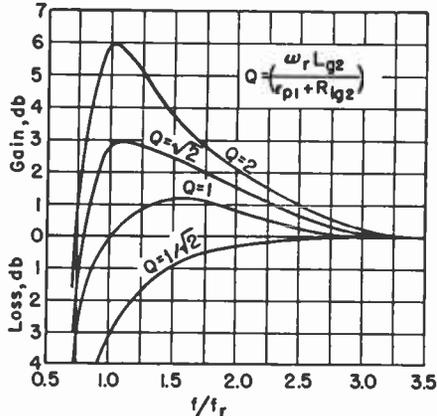


FIG. 13. Low-frequency characteristic of a double-impedance-capacitance amplifier.

At the high frequencies the amplification per stage, A_H , is

$$A_H = \frac{A_M}{\sqrt{1 + (\omega C_e' / G_e')^2}} \quad (14)$$

where $G_e' = 1/r_{p1}$ plus the conductances due to the core losses in the two chokes

$C_e = C_{pk1} + C_{op1} + C_{pk2} + C_{op2}(1 + |A_{r2}|)$ plus the effective distributed capacitances of the two chokes

The quantitative relation between C_e and G_e for different decibel losses at 10,000 cps can be obtained from the curves of Fig. 9. (See explanation in Sec. 8 for extending the ranges or finding values at another frequency.)

11. Design of Impedance-Capacitance-coupled Amplifiers. The application of the type of coupling shown in Fig. 12a to tubes of high plate resistance is limited principally by the amount of inductance that can be obtained in choke L_{o1} without a large amount of distributed capacitance. The distributed capacitance of the choke adds to the tube capacitance and, therefore, helps to lower the amplification at the high frequencies. Chokes for this purpose are sometimes wound in pie sections to reduce the distributed capacitance. Of course, for tubes having high plate resistance some of the maximum possible gain can be sacrificed by lowering R_{o2} to have a small variation in gain over the frequency range. This will make it easier to satisfy the requirements at both the highest and lowest frequency.

For tubes that have low plate resistance, the design procedure is to fix the value of R_{o2} so that it will not be greater than the maximum recommended value or the value which will keep the highest frequency response within the desired limit. The curves of Fig. 9 are useful for determining the limit to R_{o2} so far as frequency response is concerned. In this figure for this purpose G_e is equal to $g_{p1} + G_{o2}$ plus a conductance allowed for the core loss of L_{o1} . After R_{o2} is fixed, the value of L_{o1} is determined tentatively by the use of the curves in Fig. 9. For this purpose R_i on the graph becomes $r_{p1}R_{o2}/(r_{p1} + R_{o2})$. The last step is to determine C such that the loss due to it is not more than 0.25 db. In some cases it may be necessary to check the results by applying Eq. (10).

For tubes that have high plate resistance, the design procedure is about the same as the above except it may be necessary to work back and forth from h-f consideration to l-f consideration to obtain the desired characteristics.

In designing an amplifier of the type shown in Fig. 12b the general procedure is the same as above. In some cases the medium-frequency amplification may be less than μ_1 because of the core losses of the two chokes. These core losses are equivalent to two resistances in parallel from the grid to the cathode of tube 2 and their effect is similar to R_{o2} in Fig. 12a.

The following example will illustrate how to apply Eq. (13) and the curves of Fig. 13. The plate resistance r_{p1} of the tube is 10,000 ohms, the allowed resistance for R_{Lo2} is 1,000 ohms, and the desired gain at 50 cps is 3 db over the gain at medium frequencies. From the curves of Fig. 13, Q must be $\sqrt{2}$ to give the desired gain. From the expression for Q ,

$$L_{o2} = Q \frac{(r_{p1} + R_{Lo2})}{\omega_r}$$

L_{o2} is equal to $11,000/2\pi 50$ which gives 35 henrys. The size of the coupling capacitor is given by $C = 1/\omega_r^2 L_{o2}$ and is equal to 0.29 μf .

Impedance-capacitance coupling is not generally suitable for pentode tubes because of the difficulty of constructing an impedance with sufficient inductance without a large distributed capacitance. Tubes with fairly low plate resistances work best in an impedance-capacitance coupled amplifier. Such tubes have fairly low amplification factors and do not give much gain per stage as compared to the gain that can be achieved in an RC-coupled amplifier employing pentode tubes.

12. Equivalent Circuit of a Transformer-coupled Amplifier. The complete equivalent circuit of one stage of a transformer-coupled amplifier comprises the plate resist-

ance of the tube ahead of the transformer, the input capacitance of the tube after the transformer, and the equivalent circuit of the transformer itself. Figure 14 illustrates the complete equivalent circuit for one stage. This circuit does not apply to all types but represents the condition quite accurately for a great many.

In this diagram the symbols shown represent the following: μE_{g1} is the voltage generated in the tube source and r_{p1} is the plate resistance of the tube source. R_p and R_s are the primary and secondary winding resistances.

L_p and L_s are the primary and secondary leakage inductances. These inductances are due to the magnetic fluxes that link with one coil and not the other, *i.e.*, the fluxes that are not mutual to the two coils.

C_p and C_s are the effective distributed capacitances of the primary and secondary windings. C_m is the effective mutual capacitance between the windings. C_m may not be present in certain transformers. Sometimes C_m is of a complicated nature and difficult to estimate. C_L is the input capacitance of the tube load.

L_m and R_c are the magnetizing inductance and core-loss resistance of the transformer. The magnetizing current and the equivalent core-loss current of a transformer are nearly proportional to the induced voltage.

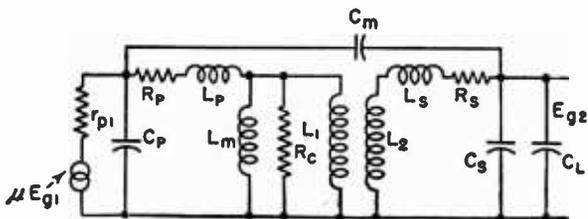


FIG. 14. Equivalent circuit of a transformer-coupled amplifier.

L_1 and L_2 are fictitious inductances necessary to transfer the current and voltage to the load and to provide the proper phase change from primary to secondary. The phase of the secondary voltage with respect to the primary is important when the mutual capacitance C_m is equal to, or greater than, 25 per cent of C_s and C_L . The ratio of the primary turns to the secondary turns is equal to $\sqrt{L_1/L_2}$. This ratio is called N , the ratio of transformation.

In Secs. 14 and 15 it is shown how the equivalent circuit is modified in order to simplify matters. This simplification is possible for a transformer which is intended to cover a range of frequencies like 50 to 5,000 cps and when the variation in amplification over the range is not more than 6 db.

13. Calculation of Transformer Constants. The material under this article applies to both interstage transformers and impedance-matching transformers. The most important constants required in a given design are the magnetizing inductance L_m and leakage inductances L_p and L_s .

The magnetizing inductance L_m is given by the expression

$$L_m \text{ in henrys} = \frac{4\pi 10^{-9} N_p^2 \mu_r A}{l} \tag{15}$$

where N_p is the number of turns on the primary and μ_r is the relative permeability; if the primary carries d.c., μ_r is the apparent incremental permeability; A is the net area of the core in square centimeters and l is the mean length of path in centimeters. When A is not the same for the entire length of the path, the total reluctance must be calculated from the sum of the reluctances of the paths over which the net area is constant. To evaluate L_m when winding carries d-c current, there must be available curves of μ_r plotted against the d-c magnetizing ampere-turns per centimeter for various flux a-c densities on the particular magnetic material.

The leakage inductances L_p and L_s depend upon the configuration of the windings.

These inductances are due to the fluxes that link with one coil and not the other. For the type illustrated by Fig. 15,

$$L_t = L_p + N^2 L_s = \frac{16\pi N_p^2}{10^9 W} \left\{ (D_1 + D_2 + 2D_o) \frac{Dl}{3} + \frac{1}{2} (D_i^2 - D_o^2) + \left[\frac{2}{3} (D_1 + D_2) + 2D_i + D_b \right] D_b \right\} \quad (16)$$

where $D_i = D_1 + D_2 + D_o$.

For an interspaced winding of this type, *i.e.*, one in which one coil is placed between the two halves of the other coil, $L_p + N^2 L_s$ is approximately one-fourth of that given by Eq. (16). All dimensions are in centimeters and are indicated in the figure.

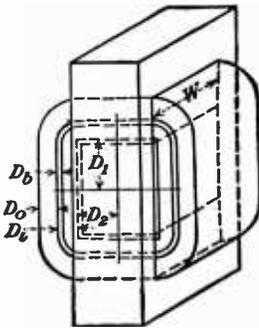


FIG. 15.

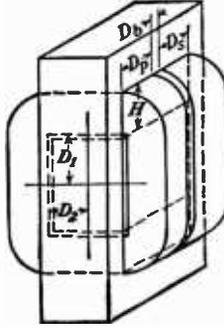


FIG. 16.

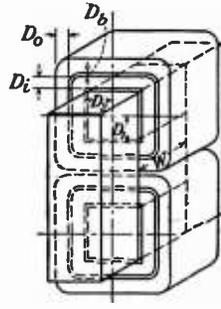


FIG. 17.

FIG. 15. Simple winding scheme for a transformer.

FIG. 16. Winding scheme for low effective capacitance.

FIG. 17. Simple winding scheme for a core-type transformer.

The leakage inductance for a winding of the type shown in Fig. 16 is approximately

$$L_t = L_p + N^2 L_s = \frac{16\pi N_p^2}{10^9 H} \left[(D_1 + D_2 + H) \left(\frac{D_p}{3} + \frac{D_s}{3} + D_b \right) \right] \quad (17)$$

For an interspaced winding of this type the total leakage inductance is approximately one-fourth of the value given by Eq. (17).

For a core-type transformer as shown in Fig. 17, in which half of each primary and secondary is wound on two opposite legs of the core, the approximate expression for the leakage inductance is

$$L_t = L_p + N^2 L_s = \frac{8\pi N_p^2}{W 10^9} \left\{ (D_1 + D_2 + D_o) \frac{Dl}{3} + \frac{1}{2} (D_i^2 - D_o^2) + \left[\frac{2}{3} (D_1 + D_2) + 2D_i + D_b \right] D_b \right\} \quad (18)$$

For an interspaced winding of the core type, *i.e.*, one in which one coil of each leg is placed between the two halves of the other coil of the same leg, the leakage inductance is approximately one-fourth the value given by Eq. (18).

For interstage and impedance-matching transformers the core losses under most ordinary circumstances are usually small compared to the copper losses, but for the sake of completeness the expression for the core-loss resistance R_c is given. It is

$$R_c = \frac{2\pi^2 10^{-16} f^2 N_p^2 A}{K_d l} \quad (19)$$

where $K_c = \frac{\text{total core loss per cc}}{B^2}$ at the operating conditions

$B =$ flux density, gauss

It is assumed that the hysteresis losses as well as the eddy-current losses are proportional to B^2 . It has been found by the author that the hysteresis losses at low flux densities are nearly proportional to B^2 , but sometimes the exponent of B is even greater than 2.

The distributed capacitance of transformer windings is due mainly to the layer-to-layer capacitances. The effective capacitance of a winding is approximately equal to the capacitance between the two mean layers divided by the number of layers. In most cases the layers may be treated as parallel plates having a dielectric equal to thickness of paper between layers plus two times the thickness of the insulation on the wire. If the dielectric constants of the paper and insulation are much different, they must be treated accordingly.

14. Theory of Transformer-coupled Amplifiers. The characteristics of this type of amplifier are best explained by dividing the frequency range into the low frequencies, the medium frequencies, and the high frequencies. The equivalent circuits of

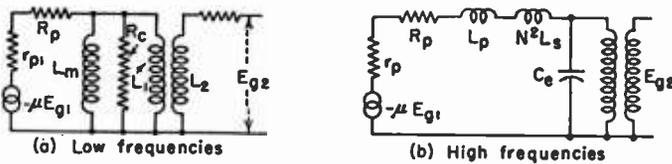


FIG. 18. Equivalent circuits of transformer-coupled amplifier.

Fig. 18a and b apply to the low and the high frequencies. At the medium frequencies the core-loss resistance R_c is usually so large compared to $r_{p1} + R_p + N^2R_c$ that the voltage amplification per stage, *i.e.*, $E_{g2}/E_{g1} = A_M$ is practically equal to μ_1/N . Hence $20 \log_{10} (\mu_1/N)$ will be used as the reference level in decibels, and the performance at the low and high frequencies will be termed a loss, or gain, in decibels measured from this reference level.

At the low frequencies the magnetizing inductance is effective and the l-f amplification A_L , in terms of A_M , is

$$A_L = \frac{A_M}{\sqrt{1 + \left[\frac{1}{\omega L_m} \frac{(R_p + r_{p1})R_c}{R_c + R_p + r_{p1}} \right]^2}} \tag{20}$$

The loss at the low frequencies due to L_m is

$$20 \log_{10} \sqrt{1 + \left[\frac{1}{\omega L_m} \frac{(R_p + r_{p1})R_c}{R_c + R_p + r_{p1}} \right]^2}$$

This case is so similar to the one illustrated by the equivalent circuit of Fig. 35 for an impedance-matching transformer that the curves given in Fig. 36 may be used to see

the relation between L_m and $\frac{(R_p + r_{p1})R_c}{R_c + r_{p1} + R_p}$ for various decibel losses at 50 cps. In many cases R_c is so large compared with r_{p1} and R_p that the quantity

$$\frac{(R_p + r_{p1})R_c}{R_c + R_p + r_{p1}} \cong R_p + r_{p1}$$

Hence in most cases $R_p + r_{p1}$ can be substituted for R_1 when using the curves in Fig. 36 to determine L_m for a given loss in decibels. The curves of Fig. 36 can be

used for any other frequency f_x by multiplying the ordinates by $50/f_x$ and locating L_m on the new scale.

The phase shift at the low frequencies is

$$\theta_L = \tan^{-1} \frac{(R_p + r_{p1})R_c}{(R_e + r_{p1} + R_p)\omega L_m} \tag{21}$$

At the high frequencies the leakage inductances and the tube and distributed capacitances affect the voltage amplification. For cases in which C_m is small compared to $C_s + C_L$, the amplification at the high frequencies in terms of A_M is

$$A_H = \frac{A_M}{\sqrt{\left(1 - \frac{f^2}{f_r^2}\right) + \frac{f^2}{f_r^2} \frac{1}{Q_r^2}}} \tag{22}$$

The gain, or loss, equals

$$20 \log_{10} \sqrt{\left(1 - \frac{f^2}{f_r^2}\right)^2 + \frac{f^2}{f_r^2} \frac{1}{Q_r^2}}$$

where $Q_r = \frac{\omega_r L_i}{R_c}$

$$\omega_r = 1/\sqrt{L_i C_s} \text{ and } f_r = 1/(2\pi \sqrt{L_i C_s})$$

$$C_s = (C_m + C_s + C_L)/N^2$$

$$R_e = r_{p1} + R_p + N^2 R_c$$

$$L_i = L_p + N^2 L_s$$

$$C_L = C_{gk2} + C_{gp2}(1 + |A_{12}|)$$

N = ratio of primary to secondary turns

The curves of Fig. 19 show how the loss, or gain, varies around the frequency f_r for different values of $\omega_r L_i/R_c$. The best results are obtained when $\omega_r L_i/R_c$ is approximately equal to 1. This can be accomplished to some extent by controlling L_i and C_s in the design.

When C_m is not small compared to $C_s + C_L$, the voltage amplification is approximately the value given by Eq. (22) times

$$1 + \frac{NC_m}{C_m + C_s + C_L} \frac{f^2}{f_r^2}$$

where N may be either positive or negative in numerical value. N is positive if the two coils form a single winding in one direction about the common core when connected together at the cathode ends, and negative when the windings are in opposite directions. The mutual capacitance may be avoided by the use of static shields.

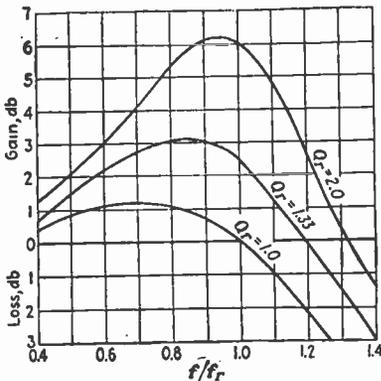


FIG. 19. High-frequency characteristics of a transformer-coupled amplifier.

The phase shift at the high frequencies is given by

$$\theta_h = 90^\circ + \tan^{-1} Q_r \left(\frac{f}{f_r} - \frac{f_r}{f} \right) \tag{23}$$

15. Design of Transformer-coupled Amplifiers. Usually transformer coupling is used with voltage amplifier tubes that have a comparatively low plate resistance. This is necessary to obtain the desirable characteristics at the low frequencies because the magnetizing inductance for a given l-f response is almost directly proportional to the plate resistance of the tube. It is essential also that the d-c plate current be as

small as possible so that it will not saturate the core of the transformer. The magnetizing inductance L_m is the first consideration in the design of an interstage transformer. The curves of Fig. 36 can be used for determining the value of L_m for a given decibel loss at the lowest frequency. In the preliminary procedure the core loss can be neglected and $R_p + r_{p1}$ can be substituted for R_1 in Fig. 36. An allowance of 8 to 10 per cent of r_{p1} is made for the primary winding resistance.

The amount of voltage amplification per stage secured at the medium frequencies is nearly equal to the amplification constant μ_1 times the ratio of secondary turns to primary turns; in the theory this is μ_1/N . Practical values for this ratio are 2 to 4. If higher, difficulty is experienced at the high frequencies because of the tube load and distributed capacitance of the secondary windings, even though the leakage inductance is very small.

The performance of the transformer at the high frequencies depends largely upon the leakage inductance and the capacitance of the secondary winding and tube load. This is illustrated in Fig. 19. For practically constant gain up to any frequency f_h either the frequency f_r must be at least two times f_h or else the winding must be so designed that $f_r = f_h$ and the quantity $\omega_r L_l / R_s = Q_r$ is approximately equal to 1.

Interspacing the windings of a transformer, placing one winding between the two halves of the other, lowers the total leakage inductance by a factor of one-fourth but generally results in a much higher effective capacitance. Therefore, the net result of interspacing is not to raise the frequency f_r by a factor of 2. Even if f_r were raised by a factor of 2, the quantity Q_r might be reduced below 1 at f_r and the gain of the amplifier would not be constant up to f_r .

Winding the transformer like Fig. 16 except with interspaced coils is very effective in reducing the capacitance of the windings, but this is very uneconomical as to space. The theory and design given here apply to input transformers as well as interstage transformers. The input transformer must be designed for a particular source impedance and a particular tube load.

16. Grid Bias and Screen Voltage.¹ The grid bias for a class A voltage amplifier is usually obtained from a resistor, R_c , in the cathode circuit as illustrated in Fig. 21 for a pentode.

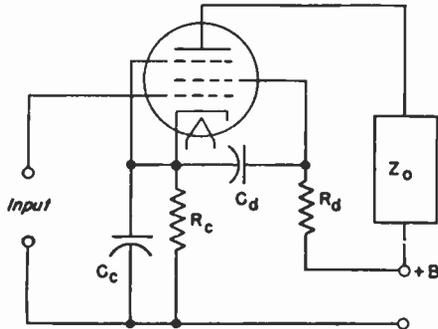
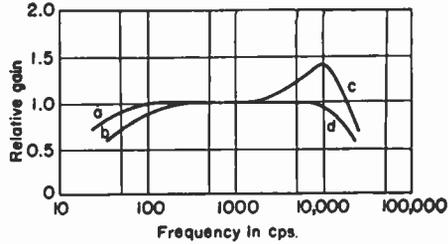


FIG. 21. Method for obtaining self-bias and screen potentials for pentode amplifier.

This is called self-bias. Each tube in a multistage amplifier is self-biased according to the d-c operating potential and plate current needed for satisfactory operation. The value of the self-biasing resistor may be computed from the relation $R_c = E_{cc} / I_b$, where E_{cc} and I_b are the d-c grid potential and plate current, respectively, needed for satisfactory operation. In case the d-c grid and plate potentials are known but not

¹ More details on the effects of self-bias and impedance in the screen circuit of a pentode amplifier are given in the following reference: F. E. Terman, W. R. Hewlett, C. W. Palmer, and W. Y. Pan, Calculation and Design of Resistance-coupled Amplifiers Using Pentode Tubes, *Trans. AIEE*, 69, 879, 1940.



a = high magnetizing inductance L_m
 b = low magnetizing inductance L_m
 c = Q_r factor, 1.41 at 10,000 cps
 d = Q_r factor, 1.0 at 10,000 cps

FIG. 20. Gain-frequency characteristic of transformer-coupled amplifiers.

the d-c plate current, the plate current may be obtained from the I_b - E_b characteristics of the tube. For pentode and beam tubes I_b is replaced by the total cathode current I_k .

The self-bias resistor must usually be shunted by a capacitor C_c except when cathode current feedback is desired (see Sec. 28) or when two tubes are operated in push-pull with a common bias resistor. For triodes which are commonly used as voltage amplifiers, the cathode bias impedance causes a gain reduction factor given by,

$$\text{Gain reduction factor caused by self-bias impedance} = \frac{1}{1 + \frac{Z_c}{r_p + Z_o}} \quad (24)$$

For pentodes,

$$\text{Gain reduction factor caused by self-bias impedance} = \frac{1}{1 + g_{p0}Z_c} \quad (25)$$

where Z_c is the impedance of the self-bias circuit and Z_o is the output impedance of the amplifier.

For both types of tubes the self-bias impedance causes a phase shift of the output voltage in a leading direction at the low frequencies. To reduce the effect on the magnitude and phase of the output voltage the susceptance of C_c should be five to ten times the conductance of R_c at the lowest frequency to be amplified. Electrolytic capacitors with 20 to 50 μf capacitance and a low voltage rating of 25 to 50 volts are usually suitable for the purpose.

When a pentode is used for a voltage amplifier, the d-c screen potential is usually obtained by connecting a dropping resistor R_d between the screen and the d-c supply for the plate circuit. The size of the dropping resistor is determined by the relation,

$$\text{Desired screen potential} = \text{d-c supply potential} - I_{c2}R_d$$

where I_{c2} is the d-c screen current.

To prevent degenerative feedback which reduces the gain, a by-pass capacitor is connected between the screen and the cathode of the tube. The susceptance C_d must be large compared to the dynamic screen conductance plus the conductance of the dropping resistor R_d at the lowest frequency to be amplified if the degenerative feedback effect is to be negligibly small. The phase shift caused by the screen circuit impedance is leading and, therefore, in the same direction as that caused by the coupling capacitor and the cathode impedance. Recommended values for the dropping resistor and by-pass capacitor are given in the RCA Receiving Tube Manual and handbook for RC-coupled amplifiers. In the absence of information on the dynamic screen conductance the size of the by-pass capacitor may be determined by experience or experiment for other applications where pentodes are used as amplifiers.

17. Power Supply to Tubes of an Amplifier. The design of the power supply is not included here. Only the things pertinent to the operation of the amplifier are given here.

Filament-power supply whether a.c. or d.c. should have good regulation. When a.c. is used, the leads should be low in resistance and twisted to avoid setting up disturbing magnetic fields.

For the B supply the importance of regulation depends upon the class of the amplifier, the class B type requiring the best regulation. It is important that the internal impedance of the supply, such as a rectifier, be small at the lowest a.f. as compared to the load impedance, particularly if the load impedance is somewhat inductive.

When a common rectifier and also low-capacity batteries for the B supply of a multistage amplifier are used, feedback will result unless means are taken to eliminate it. This feedback effect comes from a voltage setup largely by the plate current of the output tube flowing through the impedance of the B supply which is common to the plate circuits of the first stages of the amplifier. The feedback circuit is illustrated by Fig. 22. Z_c represents the impedance of the B supply which is common to all

stages of the amplifier. A is the ratio of the plate voltage of the output tube to the plate voltage of the n th tube from the output end. Z_{on} is the impedance offered to the n th tube, and r_{pn} and μ_n are the constants of the n th tube. Then the over-all voltage gain of the amplifier is

$$\text{Voltage gain of entire amplifier} = \frac{\frac{\mu Z_{on}}{r_{pn} + Z_{on}} A}{1 - \frac{Z_c A r_{pn}}{Z_{oo}(r_{pn} + Z_{on})}} \quad (26)$$

The quantity $Z_{on}A/(r_{pn} + Z_{on})$ is the voltage gain when the common impedance coupling is zero. Hence feedback from common impedance coupling changes the gain and will cause sustained oscillations when the quantity

$$\frac{Z_c A r_{pn}}{Z_{oo}(r_{pn} + Z_{on})} = 1$$

The effect of common impedance coupling can be reduced and practically eliminated by the use of simple circuits of resistance, or inductive impedance, in series and capacitance in shunt with the plate supply to each tube, as shown in Fig. 23. These are called *decoupling circuits*, and the decoupling elements are C_{d1} , C_{d2} , R_{d1} , and R_{d2} . The reactances of the decoupling capacitors should be small compared to the decoupling impedances at the lowest frequency for which the amplifier is designed. Then letting

$$D_1 = \frac{j}{\omega C_{d1} Z_{d1}} \quad \text{and} \quad D_2 = \frac{j}{\omega C_{d2} Z_{d2}} \quad (27)$$

the input tube of the amplifier of Fig. 23 will be decoupled by a factor $D = D_1 D_2$, and

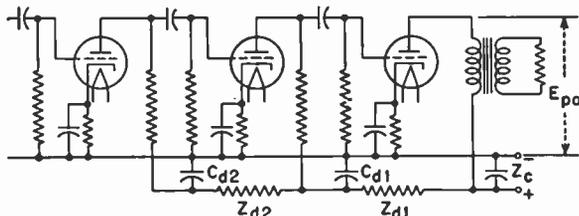


FIG. 23. Use of decoupling circuits in a three-stage amplifier. Usually Z_{d1} and Z_{d2} are resistors.

the expression for the gain of the amplifier will be

$$\frac{E_{po}}{E_o} = \frac{\text{gain with no common impedance coupling}}{1 - \frac{Z_c}{Z_{oo}} A \frac{D r_{pn}}{(r_{pn} + Z_{on})}} \quad (28)$$

where A is the gain between the plate of input tube and plate of output tube. Hence, in order substantially to eliminate the trouble from common impedance coupling, it is necessary to make $\frac{Z_c}{Z_{oo}} AD \frac{r_{pn}}{(r_{pn} + Z_{on})}$ small compared to 1. This is usually accomplished quite well by making $D = 1/A$. In the circuit shown, the middle tube is

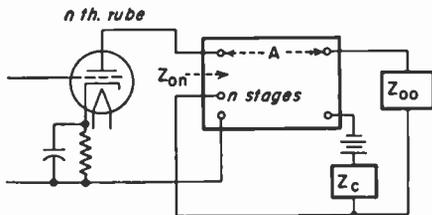


FIG. 22. Common impedance coupling between output and n th stage from output.

decoupled from both the output and input tube. This may not be necessary, but a two-section decoupling circuit is much more effective than a single decoupling section having $C_d = C_{d2} + C_{d1}$ and $Z_d = Z_{d2} + Z_{d1}$. For a filtered rectifier plate supply the common impedance Z_c is the reactance of the output filter condenser. Some decoupling can be accomplished by connecting the individual stages of the amplifier across different points of the rectifier filter. Another practice for decoupling is to use a separate decoupling resistor and capacitor for each stage instead of the cascade arrangement shown in Fig. 23.

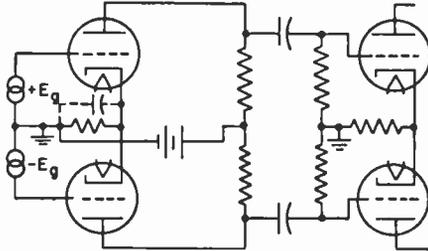


FIG. 24. Class A push-pull voltage amplifier.

18. Class A Push-pull Voltage Amplifiers. The various types of voltage amplifiers thus far described may be operated in push-pull as illustrated in Fig. 24 for an RC -coupled amplifier. The advantages of push-pull operation are as follows:

1. The hum from filament (when filament-type tubes are used and are heated from a-c) and plate supply is less.

2. Regeneration, or degeneration, from common impedance coupling in the plate supply is minimized.

3. Cathode resistors may not require by-passing capacitors. However, the practice varies and they are sometimes used.

4. Harmonic distortion is reduced.

5. For the impedance and transformer coupled amplifiers there is no d-c magnetizing flux in the cores of these components, provided both plate windings are on the same core. Because of the d-c flux cancellation the impedance or the transformer required for push-pull operation is no larger in physical dimensions than that required for single-sided operation.

The disadvantage of push-pull operation is it requires twice as many tubes, and for resistance-capacitance coupling twice as many coupling components, but gives no greater gain than a single-sided amplifier.

The design of a push-pull voltage amplifier is essentially the same as that of a single-sided amplifier. An RC -coupled amplifier may be designed as a single-sided amplifier but with twice as many components put together in a push-pull arrangement. The design of an interstage push-pull transformer does not differ materially from a singlesided transformer except for the cancellation of the d-c flux in the core and the necessity for having symmetrical leakage inductances and distributed capacitances in each half of the primary and secondary windings.

19. Phase-inverter Amplifiers.

When an amplifier requires a single-ended input, but it is desirable to have the output tubes in push-pull, it is necessary to derive voltages for the push-pull grids that are equal in magnitude and 180 deg out of phase over the complete frequency range of the amplifier. This can be done by the use of a transformer with a single primary and center-tapped secondary. However, it is somewhat difficult to design such a transformer which will have secondary voltages from each end to the center tap that are equal in magnitude and 180 deg out of phase over any considerable range of audio frequencies. Also it is often desirable to have RC coupling throughout an amplifier. This can be accomplished by the scheme of Fig. 25. Tube B is the phase-inverter tube. The input voltage to tube B is equal to the grid voltage of tube C divided by the voltage amplification of the phase-inverter stage. Usually tubes A and B are exactly alike, and stage A has the same voltage

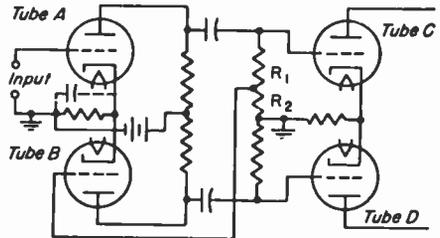


FIG. 25. Phase-inverter type of push-pull amplifier.

amplification as the phase-inverter stage. Hence,

$$R_2 = \frac{R_1 + R_2}{\text{voltage amplification of the stage}}$$

When the adjustment is made so that the grid voltages of tubes C and D are equal in magnitude and 180 deg apart at the mid-frequency of the amplifier, they will be somewhat unbalanced at both the low and high frequencies.

When the different tubes of the same type have large variations in constants, the self-balancing phase inverter of Fig. 26 is desirable. The value of R is not critical and may range from 0.1 to 0.5 of the grid resistors in the output stage.

Phase inversion may be accomplished by a single tube as shown in Fig. 27. The disadvantage of this type of phase inverter is that the voltage gain of the tube is less than 1.

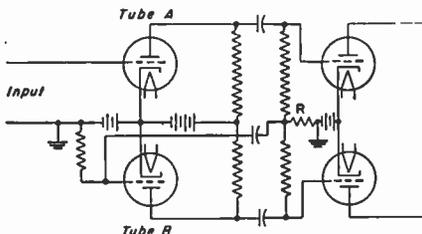


FIG. 26. Self-balancing type of phase inverter.

20. Direct-coupled Amplifiers. Under this classification are included all types of amplifiers in which the grid of one tube is connected to the plate of the preceding tube in such a manner that changes in d-c potential on the grid of the input tube will be amplified through the system. There are two important applications of such amplifiers. One application is an amplifying system for d-c purposes. The other application is an amplifier for a-c purposes when phase distortion at low frequencies is a consideration. It is difficult to obtain

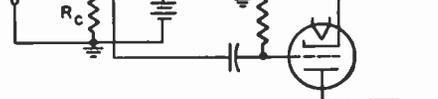


FIG. 27. Single-tube phase inverter.

much amplification at low frequencies without phase distortion by the usual types of a-c amplifiers.

Direct-coupled amplifiers have h-f characteristics like a well-designed RC amplifier. The tube capacitances shunt the coupling resistor and cause the amplification to decrease with increase in frequency above the frequency at which the effective shunt-capacitance susceptance is about three times the combined conductance of the coupling resistor and plate conductance.

The one common fault with many of the direct-coupled amplifiers when used for d-c work is instability. Small changes in the filament-, plate-, and grid-supply voltages cause false results in the output device. For amplifying l-f a.c. this particular characteristic is not so objectionable. Another common objection is the nature of the plate- and filament-supply voltages that are required.

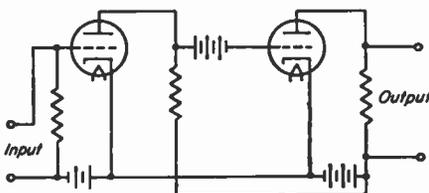


FIG. 28. Direct coupling for voltage amplifier.

Figure 28 illustrates the principles of direct coupling. A change in the d-c voltage of the grid of the first tube will result in an amplified change in the output voltage. However this amplifier is not very practical because of the separate d-c voltage supplies required and, more important, because the instability of these d-c sources causes false changes in the output voltage. The scheme illustrated by Fig. 29, known as the Loftin-White amplifier, combines all the d-c sources into a single source of fairly high

voltage for two stages of direct coupling. The Loftin-White scheme may also be used for a single-stage amplifier by omitting the last tube. It may also be arranged in push-pull by adding the same type of tubes and identical circuit elements and connecting corresponding circuits to the same point on the voltage divider. For certain applications it is sometimes desirable to connect the cathode and the screen of the second tube to a second voltage divider which is in parallel with the one shown. The lower output terminal is also connected to the second voltage divider. This arrangement reduces the reaction caused by the plate-current changes of the second tube on the

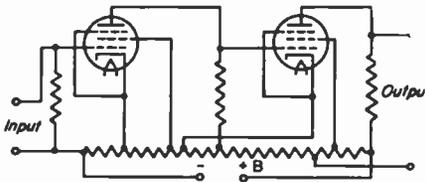


FIG. 29. Direct-coupled amplifier employing a single voltage source with potential divider.

grid of the first tube. For highest stability, particularly for the single-sided amplifier, across certain points of the voltage divider will sometimes improve the stability.

Figure 30 illustrates a type of direct-coupled amplifier that may be operated either push-pull or single-sided, provided R is equal to or greater than $10(R_{oa} + r_p)/\mu$, where r_p is the dynamic plate resistance and μ is the amplification factor of the tube. For either operation it has the characteristics of a push-pull amplifier and high stability by virtue of the degenerative feedback provided by resistor R . It is a convenient type of amplifier for a d-c electronic voltmeter. A d-c microammeter connected across the output terminals may be calibrated to read voltages in positive or negative polarity to ground by switching one grid or the other to the ungrounded side of the unknown voltage.

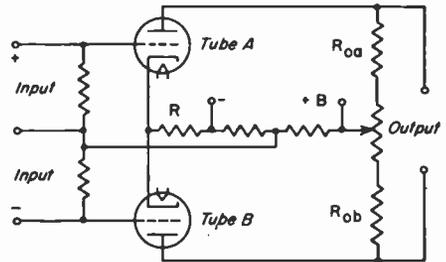


FIG. 30. Direct-coupled amplifier with either single-sided or push-pull input.

ingenious methods such as the one proposed by Schmitt¹ for obtaining practically the maximum possible voltage amplification from a high- μ pentode such as type 57 have merit. The d-c plate potential is supplied to the pentode through a similar pentode which acts as a very high a-c impedance. The arrangement is shown in Fig. 31. The full gain of the tube is obtained only by a load of very high impedance.

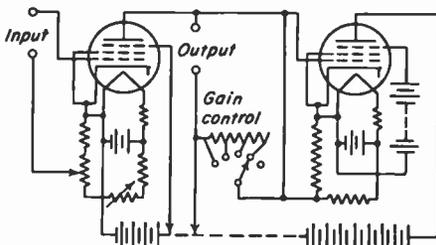


FIG. 31. Direct-coupled high-gain amplifier of Schmitt.

For this tube the voltage amplification is not a consideration, but the power sensitivity and the amount of power that can be converted without appreciable distortion are important. The power sensitivity is the power output in watts for a unit volt impressed on the grid.

¹ Schmitt, Otto H. A., A Method of Realizing the Full Amplification Factor of High- μ Tubes, *Rev. Sci. Instruments*, December, 1933.

SINGLE-TUBE POWER AMPLIFIERS

21. Class A Triode Power Amplifier.

The tube that is used to deliver power to a utilization device such as a loud-speaker is generally called a *power*

The power sensitivity is given by the expression,

$$\text{Power sensitivity} = \frac{\mu^2 R_o}{(R_o + r_p)^2} \tag{29}$$

when the output impedance is a pure resistance R_o . The power sensitivity is a maximum and equal to $\mu^2/4r_p$ when $R_o = r_p$. However, this is not the best value of R_o for maximum undistorted power output. From theoretical considerations, maximum undistorted power output is obtained when $R_o = 2r_p$ and when the peak a-c input voltage is equal to the grid-bias voltage. When $R_o = 2r_p$,

$$P_o = \frac{2\mu^2 E_g^2}{9r_p} \tag{30}$$

where E_g is the rms value of the a-c input voltage. For maximum undistorted power output, $E_g \sqrt{2}$ is equal to the grid-bias voltage. Because the current-voltage characteristics of a tube are not straight lines, the output resistance R_o should usually be greater than $2r_p$ to limit the second-harmonic current to 5 per cent of the fundamental.

The maximum power output and second-harmonic distortion can be calculated approximately for assumed values of load resistance by applying the following relations and referring to Fig. 32:

$$\text{Power output} = \frac{(I_{\max} - I_{\min}) \times (E_{\max} - E_{\min})}{8} \tag{31}$$

$$\text{Per cent second-harmonic distortion} = \frac{\frac{I_{\max} + I_{\min}}{2} - I_o}{I_{\max} - I_{\min}} \times 100 \tag{32}$$

$$\text{Efficiency of power conversion (plate-circuit efficiency)} = \frac{(I_{\max} - I_{\min})(E_{\max} - E_{\min})}{8I_{b,d}E_{b,b}} \tag{33}$$

where $I_{b,d}$ is the dynamic average value of the d-c plate current and $E_{b,b}$ is the plate-supply potential. $I_{b,d}$ is equal to the quiescent current I_o plus the peak amplitude of the second harmonic. Practical efficiencies range from 20 to 30 per cent. Recommended load resistors and d-c operating potentials for class A triode power amplifiers are generally given in the tube manuals and handbooks.

In view of the above relationship between the quiescent current and the dynamic average current, a fairly accurate value for the amplitude for the second harmonic may be obtained experimentally by measuring the difference between the dynamic average current (d.c. with input signal) and the quiescent current (d.c. with no input signal).

22. Class A Pentode- and Beam-tube Power Amplifiers. The power sensitivity and the efficiency of power conversion for pentode and beam-tube power amplifiers are usually much higher than that of triode tubes. Further points in favor of pentode and beam-tube power amplifiers are their lower input capacitances and lower driving voltages as compared to comparable triodes.

The distortion characteristics of pentode and beam tubes are similar but radically different from those of a triode. For a triode amplifier the distortion, mostly second harmonic, decreases as the load resistance is increased. It reaches a tolerable value for

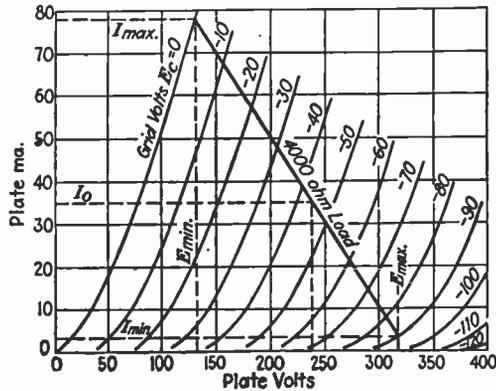


FIG. 32. Load characteristics of a triode.

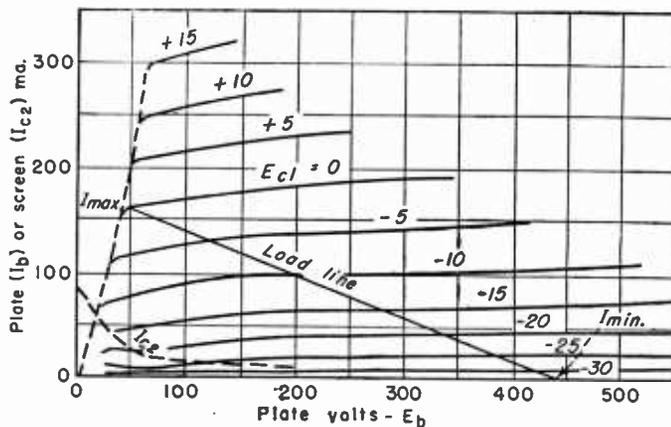


FIG. 33. Typical I_b - E_b characteristics and load line for a beam-tube class A power amplifier. $E_{c2} = 250$ volts.

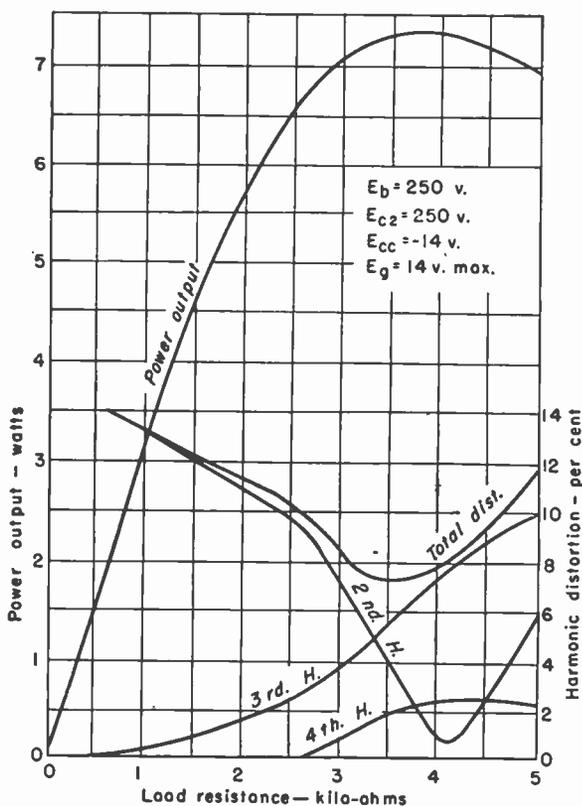


FIG. 34. Distortion and power output vs. load resistance for a beam-tube class A power amplifier.

a load resistance equal to or somewhat greater than twice the plate resistance of the tube. For pentode- and beam-tube power amplifiers there is third- and fourth-harmonic distortion in addition to second-harmonic distortion. The total distortion is a minimum for load resistances that are only a fraction of the plate resistances.

Figure 33 illustrates the mode of operation of a beam-tube amplifier with respect to its I_b - E_b characteristics. A pentode tube operates in a similar manner. Figure 34 is typical of the manner in which the power output and distortion depend on the load resistance for pentode and beam tubes. For the particular tube represented in Fig. 34 the total distortion is a minimum for a load resistance of 3,500 ohms. The plate resistance of this tube is 22,500 ohms. The recommended load resistance is 2,500 ohms. This load resistance gives a power output of 6.5 watts and total harmonic distortion of 10 per cent. Thus it is seen that load resistances for tubes of this kind are rather critical and only a fraction of the plate resistance. Consequently pentode and beam tubes should be used for single-ended class A amplifiers only when some distortion can be tolerated and when the load resistance stays within certain limits with respect to the frequency of operation or other variables unless degenerative feedback is used for reducing the distortion. Balanced push-pull operation gives good results because of the even harmonic cancellation.

The expressions for the harmonic distortion have to be extended to include not only the second but also the third and fourth harmonics for pentode and beam tubes. They are as follows:

$$I_1 = \frac{I_{\max} - I_{\min} + 1.41(I_{0.707} - I_{-0.707})}{4} \tag{34}$$

$$I_2 = \frac{2I_{b0} - (I_{\max} + I_{\min})}{4} \tag{35}$$

$$I_3 = \frac{1.41(I_{0.707} - I_{-0.707}) - (I_{\max} - I_{\min})}{4} \tag{36}$$

$$I_4 = \frac{2(I_{\max} - I_{\min}) - (I_{0.707} + I_{-0.707}) - 2I_{b0}}{4} \tag{37}$$

$$I_{bd} = I_{b0} - I_2 - I_4 \tag{38}$$

where I_{b0} = quiescent value of plate current
 $I_{0.707}$ and $I_{-0.707}$ = plate currents at instants the grid voltages are $E_{cc} + 0.707E_c$ and $E_{cc} - 0.707E_c$, respectively
 I_{bd} = dynamic average value of plate current

The d-c power input to the plate is $E_b I_{bd}$. The d-c power input to the screen should be added to that of the plate to get the total d-c power for operating the tube. The useful a-c power output is $I_1^2 R_L / 2$.

23. Impedance-matching Transformers. When a given load resistance R_L is not of the proper magnitude to result in maximum power into the load from a source which

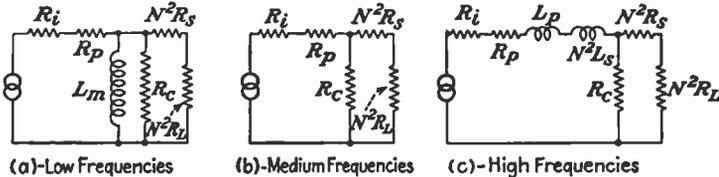


FIG. 35. Equivalent circuits of an impedance-matching transformer.

has a resistance R_i , a transformer is interposed between the source and the load. Because of the resistances of the transformer windings and the losses in the magnetic core the transformer will consume a certain amount of energy itself. In addition to the energy lost in the transformer the magnetizing current causes a loss of power to the load at the low frequencies, and the leakage inductance causes a loss at the high fre-

quencies. For a transformer of this type, intended to cover a range of frequencies, it is convenient to divide the theory and design into three phases: low-frequency, medium-frequency, and high-frequency. Figures 35a, b, and c represent the equivalent circuits that apply to each of these phases of discussion.

In the figures R_i is the internal resistance of the source; R_p and R_s are the primary and secondary winding resistances; L_p and L_s are the leakage-flux inductances of the primaries and secondaries; L_m and R_c are the magnetizing inductance and core-loss resistance; and N is the ratio of primary turns to secondary turns.

The current in the transferred load resistance at the medium frequency is used as the reference level. Referring to Fig. 35b and letting $R_1 = R_i + R_p$,

$$R_2 = (R_s + R_L)N^2 \quad R_3 = \frac{R_2}{1 + (R_2/R_c)} \quad \text{and} \quad R_4 = \frac{R_3 R_1}{R_2 + R_1}$$

$$I_M = \frac{E}{R_2(R_3 + R_1)/R_3} \tag{39}$$

In many cases R_2/R_c is so small compared to 1 that $I_M = E/(R_2 + R_1)$.

For the low frequencies Fig. 35a applies, and the current I_L in terms of I_M is

$$I_L = \frac{I_M}{\sqrt{1 + (R_4^2/\omega^2 L_m^2)}} \tag{40}$$

Then $20 \log_{10} \sqrt{1 + (R_4^2/\omega^2 L_m^2)}$ is the loss due to L_m . Figure 36 shows the relation between L_m and R_4 for various losses at a frequency of 50 cps. For any other frequency multiply the ordinates by $50/f_x$ and locate L_m on the new scale. Also, because of the linear relation between L_m and R_4 , both scales may be changed simultaneously by any factor x to provide a more convenient range for

FIG. 36. Loss at low frequency due to magnetizing inductance.

R_4 . For most cases, since R_c is several times R_2 , the quantity R_4 is equal to $R_1/(1 + R_1/R_2)$.

For the high frequencies Fig. 35c applies, and the current I_H in terms of I_M is

$$I_H = \frac{I_M}{\sqrt{1 + \frac{\omega^2 L_l^2}{(R_1 + R_2)^2}}} \tag{41}$$

Then $20 \log_{10} \sqrt{1 + \frac{\omega^2 L_l^2}{(R_1 + R_2)^2}}$ is the loss due to the leakage inductance. Figure 37 shows the relation between the total leakage inductance

$$L_l = L_p + N^2 L_s$$

and the resistance $R_1 + R_2$ for different decibel losses at 10,000 cps. For any other frequency f_x , multiply the ordinates by $10,000/f_x$, and read L_l on the new scale. Also both scales may be changed simultaneously by a factor x in order to provide a more convenient range for $R_1 + R_2$.

The procedure in designing a transformer of this kind is first to determine the size of core and number of primary turns to obtain a value of L_m which will limit the loss to a preassigned amount. In this procedure it is necessary to allow for the winding resistances R_p and R_s . The expression for L_m is given in Sec. 13. The next step is to

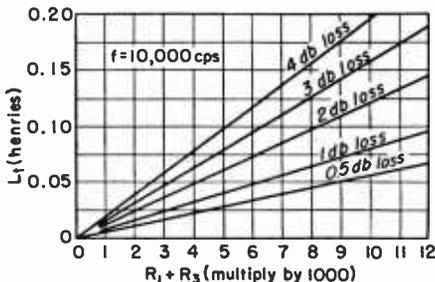
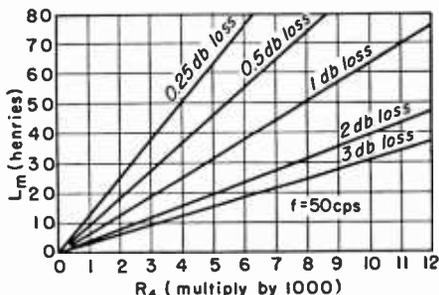


FIG. 37. Loss at high frequency due to leakage inductance.

fix the ratio of turns and the number of secondary turns for the desired value of transferred load resistance. The final step is to determine the style of winding that will keep the leakage inductance within the limit which is allowed for a given loss at the highest frequency.

PUSH-PULL POWER AMPLIFIERS: CLASS A, CLASS AB, AND CLASS B

24. Graphical Analysis for Push-pull Power Amplifiers. The circuit diagram of the push-pull type of power amplifier is shown in Fig. 38. Whenever possible, the power stage of an amplifier should be operated in push-pull. There are several advantages of push-pull operation over a single-ended power stage. When a single tube is operated so that the efficiency of power conversion is reasonably high, the harmonics are also high. In push-pull operation the even harmonics cancel out in the final load resistor. Consequently, for a given percentage of distortion, the operating voltage can be adjusted so that each tube will deliver more power into a load resistor than a similar single tube will deliver into its optimum load in a single-ended power stage. Other advantages of push-pull operation are given in Sec. 18 for push-pull voltage amplifiers. They are equally applicable for power amplifiers. In order, however, to achieve all these advantages the two tubes must be exactly alike. Unbalanced tubes in a push-pull amplifier may result in serious wave form and frequency distortion, and the push-pull system will not deliver its full power output.

The graphical analyses for all three classes of push-pull amplifiers are essentially the same. The magnetic field in the core of output transformer is a function of the algebraic sums of the a-c currents in the two sides of the primary windings. Hence the analysis is the same, except for d-c components, as though the tubes were replaced by a single class A tube which has I_b-E_b characteristics equivalent to the algebraic sum of the characteristics of the two push-pull tubes. These are called the *composite characteristics* of the push-pull unit and are illustrated in Fig. 39. Each composite curve represents the algebraic sum of the I_b-E_b curve of one tube for a grid potential of $E_c + \Delta E_g$ and the I_b-E_b curve of the other tube for a grid potential of $E_c - \Delta E_g$. E_c is the grid-bias voltage. Then the load line, which is the load resistance measured across one primary winding of the transformer, is drawn across the composite I_b-E_b curves through the d-c operating points. Current values derived from the intersection of composite load line and the composite I_b-E_b lines are the algebraic sums of the a-c currents in the two primary windings. The effect is the same as though all the a-c current flowed through one primary winding and the source impedance were equal to the reciprocal of the slope of a composite I_b-E_b curve. Hence power output is given by the relation

$$P_o = \frac{\mu^2 E_o^2 R_L N^2}{(R_L N^2 + r_p')^2} \tag{42}$$

- where E_o = rms a-c voltage from one grid to cathode
- μ = amplification factor of either tube
- r_p' = reciprocal of the slope of a composite I_b-E_b curve
- R_L = load resistance
- N = ratio of turns of one primary winding of output transformer to turns of secondary winding

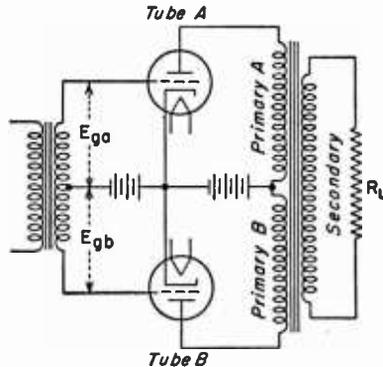


FIG. 38. Schematic of push-pull power amplifier.

The power output can also be obtained from the peak values of a-c plate current and plate voltage which are labeled I_o and E_o in Fig. 39. It is $P_o = (E_o I_o)/2$. It is seen in Fig. 39 that the composite curve and the tube characteristic for zero grid potential nearly coincide at their intersections with the load line. This is usually the case for class A operation and always so for class AB and class B. Hence it is seldom necessary to draw the composite characteristics in order to determine I_o and E_o as shown in Fig. 39. The load line is drawn with a negative slope of $1/R_L N^2$ through a point on the voltage axis equal to the d-c operating plate potential. The distortion can be obtained by plotting the current in the load resistance and analyzing the curve according to the method given in a previous section.

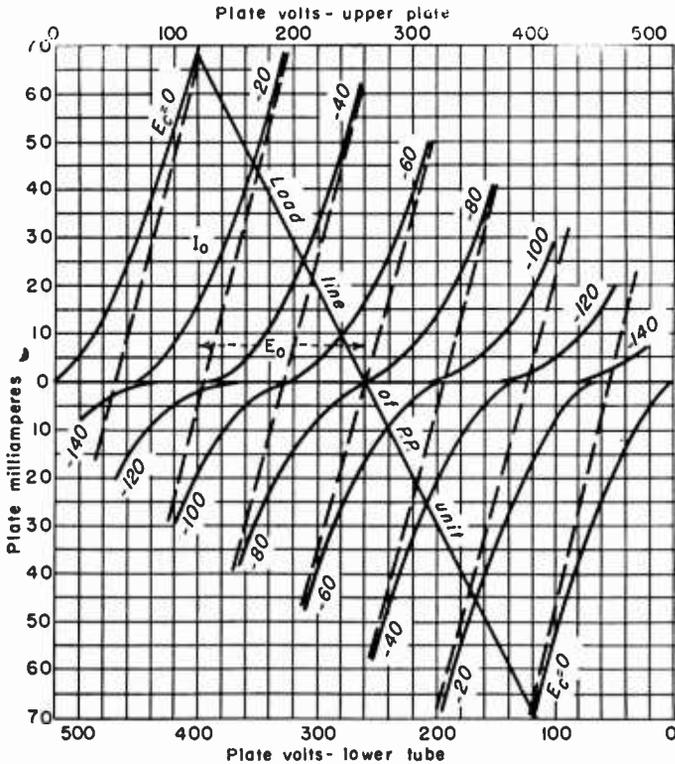


FIG. 39. Composite characteristics of push-pull amplifier.

The d-c plate current for the push-pull amplifier cannot be obtained directly from the composite characteristics. The load line of either tube must first be located. This is a line passing through points on the characteristic curve of the tube which lie directly above or below the intersections of the load line and the composite lines. The d-c component of each tube is then determined by applying the appropriate graphical harmonic analysis given in Sec. 21 or 22. This analysis will also give the harmonic plate currents. Only the odd harmonics will appear in the output.

The composite characteristics of a beam-tube push-pull power amplifier are shown in Fig. 40. The composite characteristics of a pentode are similar. It should be noted that these composite characteristics are more nonlinear than those of triodes and their slopes vary. Consequently r_p' has to be evaluated with greater care.

25. Class A Push-pull Power Amplifier. In the class A push-pull amplifier, a-c plate current flows for complete 360 deg of the input cycle, and the characteristics of the tube are nearly straight lines over the complete range of the a-c plate and grid potentials. The composite I_p - E_b curves have approximately twice the slope of the separate I_b - E_b curves from which they are derived. Hence r_p' in Eq. (42) is approximately equal to $r_p/2$ or half the plate resistance of either tube. Then for class A push-pull

$$P_o = \frac{\mu^2 E_g^2 R_L N^2}{\left(R_L N^2 + \frac{r_p}{2}\right)^2} \tag{43}$$

Equation (43) is not quite so accurate for pentode and beam tubes as for triodes because r_p' cannot be determined accurately as pointed out above.

Equation (43) furnishes information for the design of the output transformer for the class A push-pull amplifier. Referring to the design relations and curves of Sec. 22 the generator voltage is E_g , and the resistance of the source, viz., R_i , becomes $r_p/2$

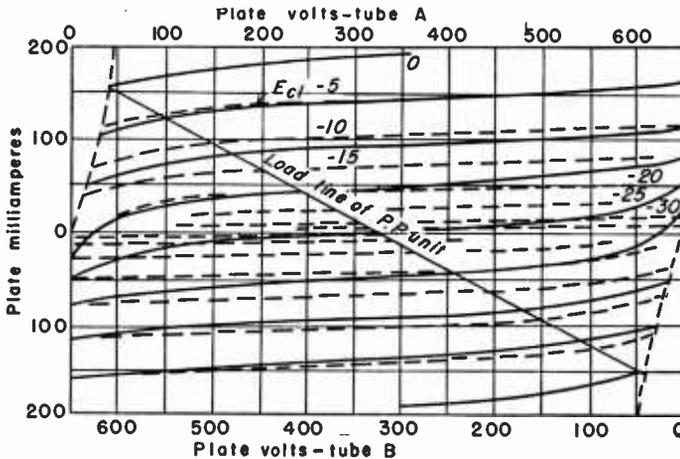


FIG. 40. Composite characteristics and load line for push-pull beam tubes in power amplifier. Dashed lines are individual tube; full lines are composite characteristics.

The transferred load resistance is $R_L N^2$, where N is the ratio of turns for one primary winding to the total secondary turns. The value of $R_L N^2$ to be used in Eq. (43) is equal to one-fourth of the plate-to-plate load resistance which is usually specified in tube handbooks as the best value to use for a given tube. Hence the allowable values for the magnetizing and leakage inductances can be determined from Figs. 36 and 37.

For pentode and beam-tube power amplifiers, since $r_p' \gg R_L N^2$, the exact value for r_p' is not so important for determining the magnetizing inductance of the output transformer.

The design of the input transformer becomes essentially the design of an interstage transformer which is given in Secs. 14 and 15. To preserve a balance in the magnitudes and phases of the two secondary voltages, the two secondary windings must be kept symmetrical with respect to leakage inductances to the primary, resistances, and distributed capacitances. If the two voltages applied to grids of the push-pull amplifier are less than 180 deg apart, the net grid-to-grid voltage which will be effective in producing output power will be less than the algebraic sum of the two voltages, and the cancellation of the even harmonic distortion will not be complete.

Class A push-pull amplifiers are usually operated with self-bias. A single resistor is connected between the common ground and the two cathodes as shown in Fig. 24. The push-pull triode amplifier usually gives more than twice the power output of a single tube amplifier because of the more favorable load resistance. The push-pull pentode or beam-tube amplifier gives approximately twice the power output of a single tube amplifier.

Class A operation gives the best wave form for the current in load resistor, but the efficiency is lower than that obtained by class AB or class B operation.

26. Class AB Push-pull Power Amplifier. In the class AB push-pull power amplifier a-c plate current of each tube flows for less than 360 deg but more than 180 deg of the input cycle. The grids may or may not be driven positive with respect to the cathode. For this type of amplifier the reciprocal of the slopes of the composite characteristics lies somewhere in between r_p and $r_p/2$, and the quantity r_p' in Eq. (42) must be determined for any particular set of operating conditions.

The design of the output transformer is carried out according to Sec. 23, where R_i becomes r_p' and $R_L N^2$ is the load impedance which must be used in establishing L_m and L_i for each primary winding. Since the two primary windings are carrying unequal currents, care must be taken in the position of each primary winding with respect to each other and to the secondary winding. The effective leakage inductance will always be higher than it would be for the same transformer if both primary windings were carrying equal and opposite currents for all points of the input cycle.

Class AB push-pull amplifiers may be so driven that their grids go positive with respect to the cathode. Hence the input transformer design depends some on whether or not there will be grid current in the secondary during a part of each positive half cycle of the grid voltage. When there is no grid current, the design is the same as that given for the input transformer of class A push-pull amplifier. When there is grid current, the load on the driver tube varies over the cycle going from no load for a part of the cycle to a maximum load current which causes quite a drop in the grid voltage of the class AB tubes. Hence the input transformer must be so designed that magnetizing inductance will be high enough for no-load conditions and have such a ratio of transformation that the output voltage of the driver tubes will not vary much over the cycle. The driver tubes should have as low plate resistance as possible. Low leakage inductance and winding resistances also help to reduce the flattening of the crest of the input voltage to class AB tubes.

Fixed or self-bias may be used for class AB power amplifiers. The increase in the d-c component of the plate current with signal voltage causes the bias voltage to increase when self-bias is used. Because of this increase of bias voltage it is necessary to operate a self-biased amplifier with a higher load resistance than is used for fixed-bias operation to keep the distortion within a reasonable limit. For triodes the power output is lower with self-bias than with fixed bias. For beam tubes the power output is about the same for either kind of bias but the distortion is slightly higher for self-bias. It is generally more satisfactory to use fixed bias or semifixed bias for beam-tube class AB₂ power amplifiers. When self-bias is used for class AB operation, the cathode resistor should be adequately by-passed. The plate supply for class AB operation should have better regulation than is needed for class A operation because of the change in d-c plate current with signal voltage.

Class AB operation gives better efficiency than class A operation with somewhat higher distortion. Because of the large even harmonics generated by each tube it is essential that the tubes be well matched at all times to achieve the optimum in even harmonic cancellation in the load resistance.

27. Class B Power Amplifiers. For class B operation the d-c grid and plate potentials are adjusted so that plate current for each tube flows for only slightly more than 180 deg of the input cycle. In the graphical construction the I_b - E_b charts for the two tubes are adjusted, for a particular set of operating voltages, so that a large part of the I_b - E_b curves of the tubes coincide with the composite I_b - E_b curves. Only the low values of I_b of each tube will be different from their algebraic sum. Hence

in this case $r_p' = r_p$ and the power output is

$$P_o = \frac{\mu^2 E_o^2 R_L N^2}{(R_L N^2 + r_p)^2} \tag{44}$$

where r_p depends somewhat upon the amplitude of E_o and should be determined for a medium value of E_o .

For class B operation each primary of the output transformer carries current of the fundamental frequency for only alternate halves of the input cycle. Hence the effective leakage inductance of the transformer is materially higher than it would be if both windings always carried equal and opposite currents. The criterion on leakage inductance is the inductance measured across only one primary winding with the secondary winding shorted. This is the leakage inductance which enters into the characteristics of the transformer at the higher frequencies. In the design procedure given in Sec. 23, r_p is the source resistance symbolized by R_i , and $R_L N^2$ is the load impedance transferred to one primary side. L_m and L_t are calculated or preassigned on the basis that only one primary winding is active at a time. The two primary windings should occupy similar positions with respect to the secondary and should be well interspaced with each other. The design of the input transformer is similar to that given for class AB operation. For zero-bias class B tubes the design of the input transformer is essentially the same as that of an impedance-matching transformer which is treated under Sec. 23.

Fixed bias is essential for class B operation. Since the d-c grid current depends upon the signal voltage and flows in a direction to unload the bias supply, it is essential to have a bias supply with good regulation. When batteries are used, they should have low internal resistance. When a rectifier is used, it should have a low output terminal impedance. Rectifiers with voltage stabilizers are quite satisfactory. For low-power output it is generally more satisfactory to use tubes in a class B amplifier that require zero bias. The plate-supply requirements for class B operation are also important because the d-c plate current is almost directly proportional to the signal voltage. Hence the plate supply should have very good regulation.

The class B power amplifier has the highest efficiency for the three types of push-pull amplifiers. Efficiencies as high as 60 per cent or better may be obtained for large power amplifiers.

FEEDBACK AMPLIFIERS

28. Principles of Degenerative Feedback. Controlled degenerative feedback is applied to a-f amplifiers for the purpose of improving their frequency characteristics, reducing wave-form distortion and phase shift, and increasing the stability. In the simplest case a voltage derived from the output of the amplifier is fed back so that it is effectively in series with the input or the grid circuit. Figure 41 illustrates degenerative feedback in its simplest form. For this simple circuit the general expression for the gain of the amplifier is

$$\text{Gain} = \frac{E_o}{E_i} = \frac{A}{1 - A\beta} \tag{45}$$

where A is the vector voltage amplification without feedback or is equal to E_o/E_g and has a negative real value, and $\beta = R_f/(R_f + R)$ when $1/\omega C \ll R$. $A\beta$ is

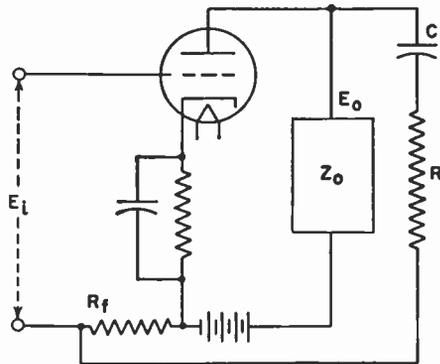


FIG. 41. Simple amplifier with output-voltage-controlled degenerative feedback.

Figure 41 illustrates degenerative feedback in its simplest form. For this simple circuit the general expression for the gain of the amplifier is

called the *feedback factor*. The performance of the amplifier as to reduction of distortion, stability, etc., depends largely on the magnitude of $A\beta$.

That feedback improves stability is shown by the following example: In the amplifier circuit shown \hat{A} has a negative numerical value. Hence the gain = $|A|/(1 + |A|\beta)$. Now assume $|A|\beta = 2$. The gain of the amplifier is equal to $|A|/3$. Suppose, owing

to a change in d-c operating conditions or the substitution of another tube of the same type, $|A|$ is increased by 20 per cent. This will result in a 6.5 per cent increase in the gain of the amplifier. Greater values for $|A|\beta$ will produce less change in gain of the amplifier. When $A\beta$ becomes large compared to 1, the gain of the amplifier is equal to $-1/\beta$ and is entirely independent of the voltage gain of the tube.

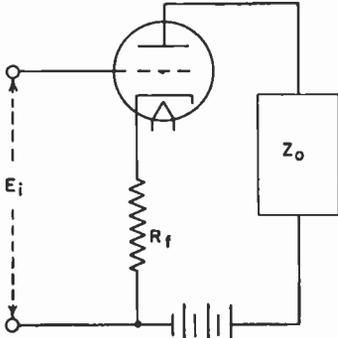


FIG. 42. Amplifier with cathode-current-controlled degenerative feedback.

Feedback as applied in Fig. 41 improves the gain vs. frequency characteristics of the amplifier. Suppose Z_o is a type of impedance that causes \hat{A} to decrease as the frequency of the input voltage is lowered. With feedback, as shown, the quantity $A\beta$ will change in the same manner with frequency. The gain with feedback will change less than \hat{A} changes with frequency. This is particularly true when $\hat{A}\beta$ is large compared to unity. This is a type of feedback that is output-

voltage controlled. Consequently the feedback stabilizes the output voltage and tends to reduce its change with frequency.

Feedback does not always reduce the change in gain with frequency. Figure 42 illustrates a type of feedback which is current controlled (cathode current feedback).

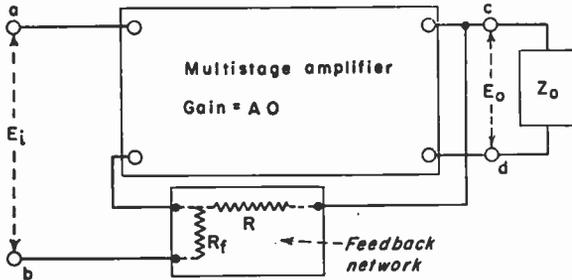


FIG. 43. Degenerative feedback applied to an entire multistage amplifier. The two resistors shown as the feedback network illustrate only one type of network that may be used. Either or both of these components may be general impedances when it is necessary to have the magnitude and angle of the feedback voltage vary with frequency.

The expression for the gain of this amplifier with feedback is

$$\text{Gain} = \frac{\hat{A}}{1 + \frac{R_f(1 + \mu)}{r_p + Z_o}} \tag{46}$$

where \hat{A} is the gain when $R_f = 0$.

Changes in gain with frequency are increased by this type of feedback. The feedback stabilizes the current through Z_o and thereby tends to cause the voltage across Z_o to change with frequency the same way that Z_o changes. Without feedback the current through Z_o changes in such a manner with frequency partly to compensate the

change in impedance and thereby cause the voltage across Z_o to change less than Z_o with frequency.

29. Feedback in Multistage Amplifiers. Either voltage- or current-controlled feedback may be applied to a multistage amplifier. It may be applied between the output and input of the amplifier or it may be applied between only a part of the amplifier. Figure 43 illustrates the application of output-voltage-controlled feedback to an entire amplifier. The feedback is output-current-controlled when the feedback voltage is proportional to the current in the impedance Z_o .

Applying the general expression for feedback, *viz.*,

$$\frac{E_o}{E_i} = \text{gain} = \frac{A}{1 - A\beta} \tag{47}$$

there results in another form

$$\text{Gain with feedback} = \frac{A}{U} / \phi - \psi \tag{48}$$

where A = magnitude of \hat{A} without feedback

ϕ = angle of A , *i.e.* the phase displacement through amplifier

$$U = \sqrt{1 - 2A \cos(\phi + \alpha)A^2\beta^2}$$

β = magnitude of $\hat{\beta}$

α = angle of $\hat{\beta}$. For the circuit of Fig. 43 when the components of the feedback network are resistances $\alpha = 0$. However feedback may be applied with a phase shift in the feed-back network

$$\psi = \tan^{-1} \frac{-A\beta \sin(\phi + \alpha)}{1 - A\beta \cos(\phi + \alpha)} \tag{49}$$

For degenerative feedback, U must be greater than unity. This condition is satisfied for all angles $(\phi + \alpha)$ from 90 to 270 deg. However when $(\phi + \alpha)$ lies between 270 and 90 deg and $A\beta$ is greater than $2 \cos(\phi + \alpha)$, there is also degenerative feedback because U is greater than unity. When $(\phi + \alpha)$ lies between 270 and 90 deg and $A\beta$ is less than $2 \cos(\phi + \alpha)$, there is regenerative feedback because U is less than unity. Oscillations will occur when $U = 0$ or when $1 + A^2\beta^2 = 2A\beta \cos(\phi + \alpha)$. Thus to predict how an amplifier with feedback will behave over a wide range of frequencies it is necessary to evaluate $A\beta$ and $\cos(\phi + \alpha)$ over the complete range. This evaluation should be carried out beyond the range of interest, both below and above, because there may be possibilities of oscillations outside the useful range. Such oscillations will render the amplifier useless.

Under certain conditions degenerative feedback reduces phase shift in an amplifier. This is easily seen for the amplifier of Fig. 43 when the phase shift α in the feedback network is zero. For effective degenerative feedback ϕ lies between 90 and 270 deg. Thus ψ is positive for ϕ greater than 180 deg and negative for ϕ less than 180 deg, and the phase shift between E_o and E_i is reduced by the angle ψ .

Figure 44 illustrates what happens to the relative gain-frequency characteristics of an RC amplifier when different amounts of feedback are applied. f_L and f_H are the lower and upper half-power frequencies without feedback. With a large amount of feedback, *i.e.*, with $A\beta$ large at the mid-frequency of the amplifier the gain below f_L and above f_H is greater than the mid-frequency gain. This is due to the phase shift

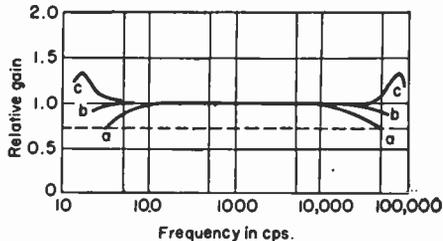


FIG. 44. Effect of feedback in RC -coupled amplifier. (a) No feedback. (b) Moderate feedback. (c) Large feedback.

through the amplifier which shifts the phase of the feedback voltage and makes the feedback less effective.

The importance of the phase angle of the feedback network should not be overlooked. In Fig. 43 when the components of the feedback are resistances as shown, the phase angle is zero. The phase angle of A , *i.e.*, the phase displacement between E_o and E_i , must lie between 90 and 270 deg for satisfactory feedback. This requires an odd number of stages in the amplifier if the amplifier is RC coupled. If α were made 180 deg by the use of an ideal transformer or if the phase of the output voltage were reversed by a transformer, an even number of stages would give the proper phase relation for satisfactory feedback. For a two-stage RC amplifier without an output transformer it is more convenient to apply the feedback as shown in Fig. 45 than to reverse the phase of the feedback voltage by the use of a transformer. This method

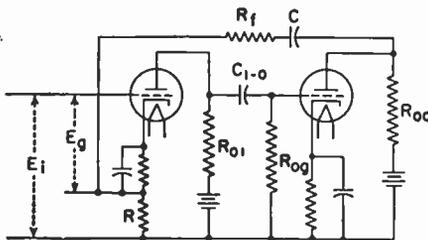


FIG. 45. Two-stage amplifier with multiple degenerative feedback.

has the same effect as reversing the feedback voltage with respect to the input voltage. This system gives some cathode current feedback in the first stage and is sometimes called *multiple feedback*.

tained by phasing the transformer voltages.

For transformer-coupled amplifiers, or if there is at least one transformer in a multistage amplifier, feedback may be applied according to Fig. 43 or 45 for either an even or an odd number of stages in the amplifier. The proper phase of the feedback voltage with respect to the input voltage can be obtained by phasing the transformer voltages. However feedback must be applied with care to a multistage amplifier which contains one or more transformers because of the large amount of phase shift caused by a transformer near its resonant frequency.

30. **Reduction in Distortion and Noise by Feedback.** Degenerative feedback reduces the nonlinear distortion generated within that portion of an amplifier enclosed by the feedback connection. This would not include any distortion existing in the input voltage which might be caused by previous equipment or by grid current of the amplifier. When the output voltage is the same with and without feedback, the reduction in nonlinear distortion is

$$\text{Distortion output (with feedback)} = \frac{\text{distortion without feedback}}{1 - A\beta} \quad (50)$$

When the output voltage of the amplifier is held constant by raising the input voltage when feedback is introduced, there is a reduction in noise and hum voltage by the same factor as given for the reduction in nonlinear distortion. Feedback is, therefore, very effective in reducing hum from the power supply which is introduced into the output stage of an amplifier.

31. **Output Terminal Impedance of a Feedback Amplifier.** Referring to Fig. 43, let the input terminals a and b be short-circuited and the load impedance be removed from terminals c and d . Then let Z_o be the impedance measured across terminals c and d without feedback and Z_o' be the corresponding impedance with feedback. The relation between Z_o and Z_o' is

$$Z_o' = \frac{Z_o}{1 - A\beta} \quad (51)$$

When the feedback is degenerative, Z_o' is less than Z_o .

When the feedback of the amplifier of Fig. 43 is controlled by the load current through a resistor R_f , the output terminal impedance Z_o'' is related to Z_o as follows:

$$Z_o'' = Z_o + R_f(1 - A) \quad (52)$$

For degenerative feedback Z_o'' is greater than Z_o .

For the simple feedback circuit of Fig. 41,

$$Z_o' = \frac{r_p}{1 + A\mu\beta} \quad (53)$$

For the simple cathode-current feedback circuit of Fig. 42,

$$Z_o'' = r_p + \mu R_f \quad (54)$$

33. Feedback Circuits. Figures 45 to 48 illustrate four ways in which feedback may be applied to amplifiers. There are other circuits and also some variations to the circuits shown. For example, in Fig. 46 the feedback circuits may be connected to the load impedance when this impedance is high and balanced to ground as would be the case if the amplifier shown were used to drive a push-pull power amplifier. In Fig. 47 the feedback voltage may be injected into the cathode circuit of the input tube and a large capacitor placed in series with the feedback connection to prevent d-c current in the load impedance. In Fig. 48 when the load impedance is coupled to the tube by means of a transformer, the feedback voltage may be injected into the cathode circuit of the input tube.

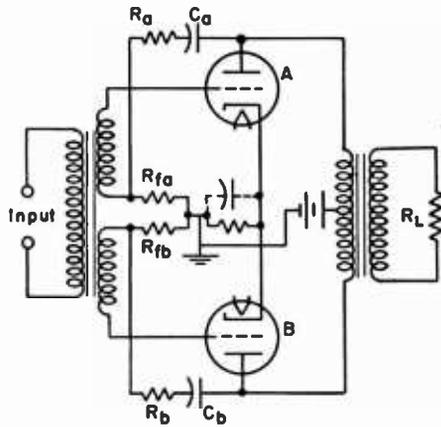


FIG. 46. Voltage-controlled feedback applied to a power amplifier. C_a and C_b may be electrolytic capacitors.

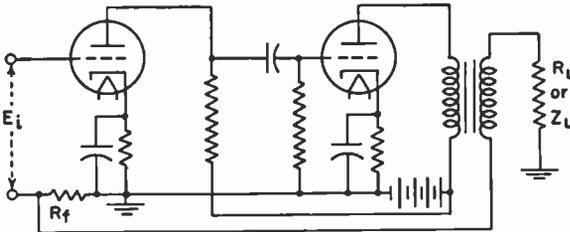


FIG. 47. Feedback controlled by the load current.

primary voltage of the transformer, the l-f response characteristics are improved, but the h-f response characteristics are poorer.

33. Cathode Follower Amplifier. The cathode follower amplifier is illustrated by Fig. 49. This is a type of degenerative feedback amplifier in which all of the output

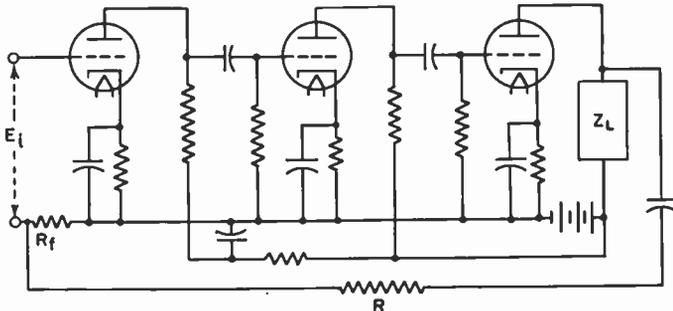


FIG. 48. Feedback applied to a three-stage RC amplifier.

voltage is in phase opposition to the input voltage. It has a very low input admittance at audio frequencies, given by the expression

$$\dot{Y}_i = j\omega[C_{gp} + C_{gk}(1 - A)] \quad (55)$$

and a high output terminal admittance which is given by the expression

$$\dot{Y}_o = \frac{1}{R_c} + \frac{\mu + 1}{r_p} + j\omega C_k \quad (56)$$

where C_{gp} and C_{gk} are the grid-to-plate and the grid-to-cathode capacitances, respectively, and C_k is the sum of all the interelectrode capacitances effectively in parallel with R_c . The voltage gain is given by the expression

$$A = \frac{\mu R_c}{R_c(1 + \mu) + r_p} \quad (57)$$

Thus the voltage gain is less than unity. When R_c is a general vector impedance \dot{Z}_c , Eqs. (56) and (57) are valid provided R_c is replaced by \dot{Z}_c .

This type of feedback amplifier is used principally for coupling a source of very high impedance to a much lower impedance load. For this purpose it takes the place of a transformer but has much better response-frequency characteristics.

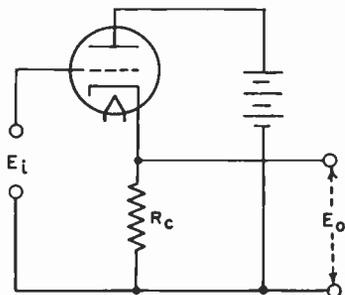


FIG. 49. Cathode follower amplifier.

GAIN CONTROL, EQUALIZATION, AND MEASUREMENTS

34. Gain Controls for Amplifiers. The gain control for an amplifier in its simplest form is a manually operated high-resistance potentiometer which is generally located in the grid circuit of one of the amplifier stages. Potentiometers for this purpose are usually tapered in resistance, or partially so, so that the gain of the amplifier will change logarithmically with the position of the potentiometer arm. Sometimes a stepped potentiometer is used which has fixed decibel changes for each step. A gain control in the grid circuit of an RC-coupled amplifier takes the place of the usual grid resistor. When a high-resistance potentiometer is placed across the secondary of a transformer for gain control, the frequency characteristics of the amplifier may be changed unless the transformer is designed especially for the purpose.

A gain control in an amplifier usually functions best in the medium level portion of the amplifier, just ahead of the first tube that may be overloaded when the input signal is highest. Any noise or hum developed in the amplifier between the gain control and the output will not be affected by the gain control. Hence in those applications for which the output may be reduced to a very low level by the gain control, it is best to have one or more stages of amplification preceding the gain control. This is particularly true for high-gain amplifiers.

Automatic gain controls known as *compressors* and *expanders* are used for certain amplifier applications. These are applications for which it is desirable to change the gain with signal level in a prescribed manner. The function of a compressor is to reduce the gain of the amplifying system for signals of the highest level. This has about the same effect as increasing the gain for the signals of lower level. It is used

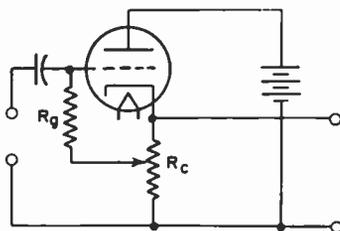


FIG. 50. Cathode follower amplifier with grid bias less than the voltage drop across cathode resistor. This gives a more favorable grid bias when R_c is very large. The input admittance is larger than that for the amplifier of Fig. 49.

for a phonograph recording amplifier and for a program amplifier in a radio-broadcast system to increase the signal-to-noise ratio for signals of low level without excessively high signal amplitudes. The expander functions in an opposite manner and amplifies the signals of higher level more than those of lower level. It is used to restore the natural ratio between signal levels which have been changed by a compressor.

An a.g.c. must function with negligible wave-form distortion. Thus the control should respond to the instantaneous average level of the over-all signal and not to each individual cycle. The circuits for a compressor and an expander are quite similar. The natural signal, after some amplification, is applied to a rectifier tube which generates a d-c output voltage that is proportional to the average signal level. This voltage is used, with proper polarity, for a part of the bias of a signal amplifier tube which has a remote cutoff. Consequently the gain of this tube decreases with increase in average signal level for a compressor and vice versa for an expander.

35. Frequency-response Control and Equalization in Amplifier Systems. By the use of certain expedients it is possible to design multistage amplifiers which will work with certain kinds of input and output devices and give over-all frequency-response characteristics of a desired type. Much can be done along this line when phase distortion is not a consideration. It may not always be desirable that the entire amplifier or each stage thereof have a response which is constant over the entire frequency band which is transmitted between the source and the load. One or more stages of transformer coupling of proper design can be used to accentuate the gain at the high frequencies. This can be done by making the Q , factors of the transformers large and make their resonant frequencies fall in the proper range. Other methods of accentuating the gain at the high frequencies can be accomplished by lowering the gain at the low frequencies. Shunting a portion of the plate resistor in an RC -coupled amplifier with inductance will lower the gain at the low frequencies. The gain of an amplifier of the l-f end of the range can be accentuated by the use of one or more stages of the double-impedance coupling which is described in Sec. 10. A satisfactory method for extending the l-f range of an RC -coupled amplifier without accentuation is a parallel circuit of resistance and capacitance placed in series with the plate coupling resistor. This expedient also compensates for the phase shift caused by the coupling capacitor and the screen and cathode by-pass capacitor. This is called l-f compensation.

With most of the standard coupling methods, such as transformer, resistance-capacitance, and impedance-capacitance coupling, it is not so easy to control the gain at the medium frequencies without effect on the gain at the low or high frequencies. In other words, the medium-frequency gain can be made greater or less than the gain at the low and high frequencies only by designing for lower or higher gain at low and high frequencies. A series circuit of resistance, inductance, and capacitance connected between the grid and cathode of one or more stages can be used to lower the gain over a small range of frequencies in the medium-frequency range. For such an arrangement the reduction in gain at the resonant frequency of the circuit depends upon the resistance of the circuit, and the band of frequencies over which the gain is reduced will depend largely upon the total effective resistance which includes the plate resistance of the tube immediately preceding the series circuit.

Variable gain control for the high frequencies, which is commonly known as *tone control*, is accomplished in its simplest manner by the use of a variable resistor and a fixed capacitance in series, both of which are placed in shunt with the coupling element of one stage of the amplifier. In a similar manner a variable resistor and a fixed inductance in series will serve as a gain control for the lower frequencies. A tone-compensated gain control is one which reduces the gain of the amplifier at the high frequencies more than at the low frequencies as the control is moved in the direction for gain reduction. This may be accomplished by a series combination of resistance and capacitance placed across a portion of the gain-control potentiometer.

There are so many combinations of methods which may be employed to give frequency-response equalization in amplifier systems and to give any desired frequency

response that it is impossible to cover all of them. Among these are the use of low-pass, high-pass, and band-pass filter circuits which are treated in another section of this handbook.

36. Frequency-response Measurements. A universal arrangement of equipment for making gain, or loss, measurements over a range of frequencies is shown in Fig. 51. The method is simply one of measuring the ratio of the output voltage to the input voltage. A calibrated potential divider or two calibrated resistors R_1 and R_2 , so arranged that R_1 plus R_2 is constant, facilitates in making these measurements.

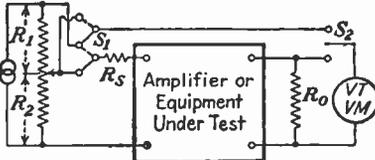


Fig. 51. Method for making frequency-response measurements.

For making gain measurements, S_1 is thrown in the position indicated by the full lines; for loss measurements in the dotted-line position. When the divider is so adjusted that the reading of the vacuum-tube voltmeter is the same for the two positions of S_2 ,

$$\text{Gain, or loss, in db} = 20 \log_{10} \frac{R_1 + R_2}{R_2}$$

For the full-line position of S_1 the resistance R_2 must always be small compared to the input impedance of the equipment under test plus R_1 . To get the true gain, or loss, characteristic of a piece of equipment as it is actually used, it must be terminated as used and the termination included in the test. For example, R_i and R_o represent the input and output resistance of the amplifier under test. These may also be any kind of impedances. When testing an input or interstage transformer it should be terminated in the tubes for which it is intended. Care must also be taken to limit the voltage applied to the equipment to the proper value.

Another method for making frequency-response measurements (particularly for high-gain amplifiers) uses one or two vacuum-tube voltmeters and a calibrated attenuator. The attenuator (usually calibrated in decibels or microvolts) is connected between a constant-voltage source and the input of the amplifier. The output voltage of the amplifier is kept constant by adjusting the attenuator as the frequency is changed. A less refined method uses a single high-grade multi-range vacuum-tube voltmeter to read the input and output voltages directly.

37. Measuring Distortion in Amplifiers. The simplest method for measuring the total harmonic distortion in the voltage across the output impedance in a power amplifier is shown in Fig. 52. For a given voltage impressed upon the grid of the amplifier, the vacuum-tube voltmeter is made to read a minimum by adjusting slide C of R_2 and the mutual inductance M . Then the reading of a vacuum-tube meter is a measure of the square root of sums of the squares of all the harmonic voltages across R_o . Mutual inductance M provides for a phase shift from 180 deg through the amplifier tube. The vacuum-tube voltmeter must be as nearly an rms meter as possible. The source should be reasonably free from harmonics. Switch S provides for measuring the total a-c voltage across R_o when a vacuum-tube meter has a multiplier to extend its range. A voltmeter, using a type 56 or 76 tube and operated over a region in which the square root of the plate current plotted against grid voltage is nearly a straight line, makes an excellent meter for this purpose.

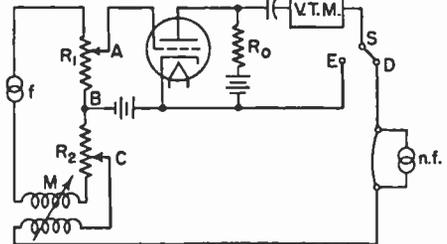


Fig. 52. Circuit for measuring distortion.

When it is desired to know the separate harmonics in the output impedance, a voltage having a frequency nf almost equal to the harmonic sought may be introduced into the connection up to contact d , as illustrated. The voltage of nf will be equal to

the particular harmonic voltage when the swing of the needle of a vacuum-tube meter is a maximum. The measurements may be carried out by means of a laboratory oscillator for nf and some filtering for the voltage obtained from the 60-cycle lighting circuit for f .

For the more refined measurements of distortion there are various types of wave analyzers on the market. These have a wider range of application than the simple method described above.

38. Measuring the Impedance of a Transformer and an Iron-core Reactance.

One of the simplest methods for measuring the impedance of an iron-core reactance at low frequencies and preferably the power frequency of 60 cycles is illustrated in Fig. 53. The circuit is arranged, when necessary, so that d.c. can be sent through the iron-core coil. When R_s is so adjusted that the reading of the vacuum-tube voltmeter is the same for both positions, a and b , of switch S , the absolute value of the impedance Z_x is equal to R_s , provided R_s is at least 20 times R . The error is less than 5 per cent. Often it is necessary to use an amplifier ahead of the vacuum-tube voltmeter. It is essential that the vacuum-tube voltmeter or amplifier be connected as shown, or false readings may result if the meter places too much stray shunt capacitance across Z_x .

The method of Fig. 53 may be used for measuring the impedances of the primary and secondary of a transformer. It is not possible, of course, to obtain the resistance and reactance separately by this method. Methods that place the standard resistance R_s in series with Z_x and require balancing the voltage drop across R_s against that across Z_x for the same current are objectionable except for quite low values of impedance. By such a method the d.c. through and a-c potential across Z_x are disturbed while adjusting R_s .

It is not generally safe to use the method described to measure the leakage inductance of a transformer. Leakage inductance is measured by shorting the secondary and measuring the impedance of the primary. Generally this measurement requires an inductance bridge because of the high value of R compared with X .

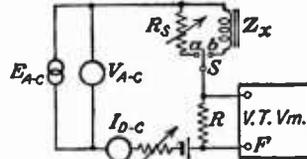


FIG. 53. Circuit for measuring impedance of iron-core coils.

39. Resistance-coupled Amplifier Chart (from the RCA Receiving Tube Manual).
 (See tables on pages 383 to 389.)

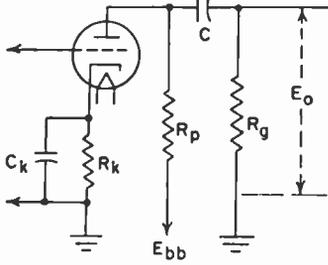


DIAGRAM 1. Triode amplifier. C and C_k chosen to give 100-cycle output voltage = $0.8E_o$. For any other value of f , multiply values of C and C_k by $100/f$. C_k value given for d-c operation; increase it if necessary to reduce hum in a-c operation. It may be necessary to operate heater at a positive voltage of 15 to 40 with respect to cathode. For any value of R_p the upper frequency at which the response falls off is above the audio-frequency range.

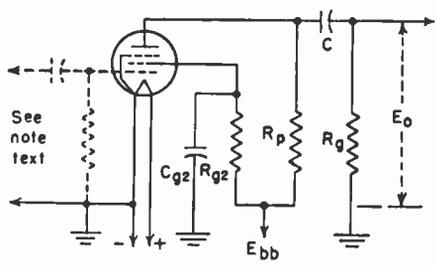


DIAGRAM 2. Pentode (filament-type) amplifier. For C and C_{g2} see note under Diagram 1 for C and C_k . For R_p of 0.1, 0.25, and 0.5 megohm the upper frequency at which the response begins to fall off is 20,000, 10,000, and 5,000 cycles, respectively. Product of the input capacitance in microfarads and grid resistor in megohms should be between 0.02 and 0.1. Values commonly used are $0.005 \mu f$ and 10 megohms.

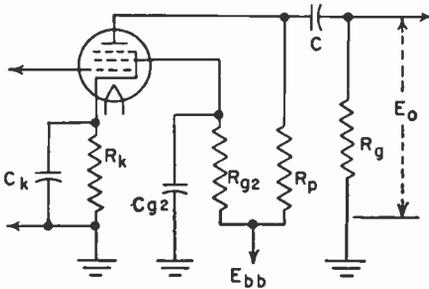


DIAGRAM 3. Pentode (heater-cathode-type) amplifier. C , C_k , and C_{g2} chosen to give 100-cycle output voltage = $0.7E_o$. For other values of f and for a-c operation see notes under Diagram 1.

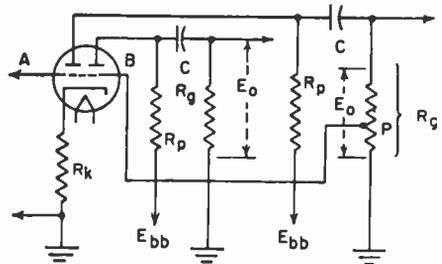


DIAGRAM 4. Phase inverter. C chosen to give 100-cycle output voltage = $0.9E_o$. For any other value of f multiply values of C by $100/f$. Tap P chosen to make voltage output of unit B equal to that of unit A . Location is given by the voltage-gain values given in charts to follow. If V.G. from chart is 20, P is located to supply $1/20$ of the voltage across R_o to the grid of unit B . Cathode capacitor may be omitted unless required to minimize hum; omitting it aids in balancing the output stages. R_k values in charts are given on basis that both units are operating at the same values of plate load and plate voltage.

Symbols Used in RC-coupled Amplifier Charts

E_{bb} = plate-supply voltage, volts
 R_p = plate resistor, megohms
 R_g = grid resistor, megohms
 R_{g2} = screen resistor, megohms
 R_k = cathode resistor, ohms

C_{g2} = screen by-pass capacitor, μ f
 C_k = cathode by-pass capacitor, μ f
 C = blocking capacitor, μ f
 E_o = peak output voltage, volts
 V.G. = voltage gain at 5 volts (rms) output unless otherwise specified

1L4 (see Circuit Diagram 2):

1S5, 1U5 (see Circuit Diagram 2):

E_{bb}	R_p	R_g	R_{g2}	R_k	C_{g2}	C_k	C	E_o	V.G.	E_{bb}	R_p	R_g	R_{g2}	R_k	C_{g2}	C_k	C	E_o	V.G.
45	0.22	0.22	0.24	—	0.071	—	0.011	12	16*	45	0.22	0.22	0.26	—	0.042	—	0.013	14	17
		0.47	0.32	—	0.06	—	0.006	14	23			0.47	0.36	—	0.035	—	0.006	17	24
		1.0	0.39	—	0.056	—	0.0035	18	30			1.0	0.4	—	0.034	—	0.004	18	28
	0.47	0.47	0.57	—	0.049	—	0.0052	14	22		0.47	0.82	—	0.025	—	0.0055	14	25	
		1.0	0.64	—	0.047	—	0.0035	17	30		1.0	1.0	—	0.023	—	0.003	17	33	
		2.2	0.74	—	0.044	—	0.0018	19	33		2.2	1.1	—	0.022	—	0.002	18	38	
1.0	1.0	1.1	—	0.036	—	0.0028	14	28	1.0	1.0	1.9	—	0.019	—	0.003	14	31		
	2.2	1.25	—	0.035	—	0.0018	16	32	2.2	2.0	—	0.019	—	0.002	17	38			
	3.3	1.45	—	0.032	—	0.0015	18	38	3.3	2.2	—	0.018	—	0.0015	18	43			
90	0.22	0.22	0.4	—	0.089	—	0.011	26	28	90	0.22	0.22	0.5	—	0.05	—	0.011	31	25
		0.47	0.46	—	0.081	—	0.0055	36	36			0.47	0.59	—	0.05	—	0.006	37	34
		1.0	0.47	—	0.08	—	0.0035	42	41			1.0	0.67	—	0.042	—	0.003	40	41
	0.47	0.47	0.84	—	0.07	—	0.0055	30	34		0.47	1.2	—	0.035	—	0.005	31	37	
		1.0	0.9	—	0.069	—	0.003	38	42		1.0	1.4	—	0.034	—	0.003	36	47	
		2.2	1.0	—	0.062	—	0.0018	40	50		2.2	1.6	—	0.031	—	0.002	40	57	
1.0	1.0	2.0	—	0.045	—	0.0028	30	45	1.0	2.5	—	0.026	—	0.003	31	45			
	2.2	2.1	—	0.045	—	0.0018	35	55	2.2	2.9	—	0.025	—	0.002	36	58			
	3.3	2.2	—	0.044	—	0.0012	40	61	3.3	3.1	—	0.024	—	0.0012	38	66			
135	0.22	0.22	0.5	—	0.09	—	0.011	42	34	135	0.22	0.22	0.66	—	0.052	—	0.011	45	31
		0.47	0.63	—	0.074	—	0.0055	54	51			0.47	0.71	—	0.051	—	0.006	56	41
		1.0	0.67	—	0.072	—	0.0035	57	60			1.0	0.86	—	0.039	—	0.003	60	54
	0.47	0.47	1.1	—	0.071	—	0.005	47	49		0.47	1.45	—	0.042	—	0.005	46	44	
		1.0	1.4	—	0.06	—	0.0028	54	68		1.0	1.8	—	0.034	—	0.003	54	62	
		2.2	1.5	—	0.051	—	0.0018	60	87		2.2	1.9	—	0.033	—	0.002	60	71	
1.0	1.0	2.1	—	0.059	—	0.0025	45	53	1.0	3.1	—	0.03	—	0.003	45	56			
	2.2	2.4	—	0.054	—	0.0018	57	88	2.2	3.7	—	0.029	—	0.0015	53	76			
	3.3	2.7	—	0.049	—	0.0012	61	91	3.3	4.3	—	0.026	—	0.0014	56	88			

* At 4 volts (rms) output.

1U4 (see Circuit Diagram 2):

2A6, 6B6-G, 6SQ7, 6SQ7-GT, 12SQ7, 12SQ7-GT, 75 (see Circuit Diagram 1):

E_{bb}	R_p	R_g	R_{g2}	R_k	C_{g2}	C_k	C	E_o	V.G.	E_{bb}	R_p	R_g	R_{g2}	R_k	C_{g2}	C_k	C	E_o	V.G.
45	0.22	0.22	0.06	—	0.048	—	0.011	11	23	90	0.1	0.1	—	6,300	—	2.2	0.02	3	23*
		0.47	0.07	—	0.045	—	0.006	15	33			0.25	—	6,600	—	1.7	0.01	5	29†
		1.0	0.11	—	0.04	—	0.003	17	39			0.5	—	6,700	—	1.7	0.006	6	31‡
	0.47	0.47	0.34	—	0.025	—	0.005	13	34		0.25	—	10,000	—	1.24	0.01	5	34†	
		1.0	0.44	—	0.022	—	0.003	16	46		0.5	—	11,000	—	1.07	0.006	7	40‡	
		2.2	0.5	—	0.022	—	0.002	18	55		1.0	—	11,500	—	0.9	0.003	10	40	
1.0	1.0	1.0	—	0.016	—	0.003	14	43	0.5	—	16,200	—	0.75	0.005	7	39			
	2.2	1.0	—	0.016	—	0.002	17	51	1.0	—	16,600	—	0.7	0.003	10	44			
	3.3	1.1	—	0.015	—	0.001	17	60	2.0	—	17,400	—	0.65	0.0015	13	48			
90	0.22	0.22	0.3	—	0.046	—	0.01	27	37	180	0.1	0.1	—	2,600	—	3.3	0.025	16	29
		0.47	0.36	—	0.04	—	0.006	36	54			0.25	—	2,900	—	2.9	0.015	22	36
		1.0	0.4	—	0.038	—	0.003	39	63			0.5	—	3,000	—	2.7	0.007	23	37
	0.47	0.47	0.9	—	0.027	—	0.0045	29	61		0.25	—	4,300	—	2.1	0.015	21	43	
		1.0	1.0	—	0.023	—	0.003	35	82		0.5	—	4,800	—	1.8	0.007	28	50	
		2.2	1.1	—	0.022	—	0.002	38	96		1.0	—	5,300	—	1.5	0.004	33	53	
1.0	1.0	1.9	—	0.02	—	0.0025	30	77	0.5	—	7,000	—	1.3	0.007	25	62			
	2.2	2.0	—	0.02	—	0.002	35	98	1.0	—	8,000	—	1.1	0.004	33	57			
	3.3	2.2	—	0.018	—	0.001	37	114	2.0	—	8,800	—	0.9	0.002	39	58			
135	0.22	0.22	0.4	—	0.052	—	0.011	44	46	300	0.1	0.1	—	1,900	—	4.0	0.03	31	31
		0.47	0.49	—	0.037	—	0.005	55	71			0.25	—	2,300	—	3.5	0.015	41	39
		1.0	0.52	—	0.034	—	0.003	60	83			0.5	—	2,300	—	3.0	0.007	45	42
	0.47	0.47	1.1	—	0.029	—	0.0045	45	77		0.25	—	3,300	—	2.7	0.015	42	48	
		1.0	1.3	—	0.023	—	0.003	53	106		0.5	—	3,900	—	2.0	0.007	51	53	
		2.2	1.4	—	0.022	—	0.002	59	123		1.0	—	4,200	—	1.8	0.004	60	56	
1.0	2.2	2.3	—	0.021	—	0.0025	45	104	0.5	—	5,300	—	1.6	0.007	47	58			
	2.2	2.5	—	0.019	—	0.0015	53	136	1.0	—	6,100	—	1.3	0.004	62	60			
	3.3	2.9	—	0.016	—	0.001	56	163	2.0	—	7,000	—	1.2	0.002	67	63			

* At 2 volts (rms) output. † at 3 volts (rms) output. ‡ at 4 volts (rms) output.

RC-coupled Amplifier Charts. (Continued)

2B7, 6B7, 6B8, 6B8-G, 12C8 (see Circuit Diagram 3); 6A6*, 6N7*, 6N7-GT*, 63* (see Circuit Diagram 4):

E _{bb}	3)										E _{bb}	4)									
	R _p	R _g	R _{g2}	R _k	C _{g2}	C _k	C	E _o	V.G.	R _p		R _g	R _{g2}	R _k	C _{g2}	C _k	C	E _o	V.G.		
90	0.1	0.1	0.37	2,000	0.07	3.0	0.02	19	24	0.1	0.1	—	1,900†	—	—	0.025	13	16			
		0.25	0.5	2,200	0.07	3.0	0.01	28	33		0.25	—	2,250†	—	—	0.015	19	19			
		0.5	0.6	2,000	0.06	2.8	0.006	29	37		0.5	—	2,500†	—	—	0.006	20	20			
	0.25	0.25	1.18	3,500	0.04	1.9	0.008	26	43	0.25	0.25	—	4,050†	—	—	0.01	16	20			
		0.5	1.1	3,500	0.04	2.1	0.007	33	55		0.5	—	4,950†	—	—	0.006	20	22			
		1.0	1.35	3,500	0.04	1.9	0.003	32	65		1.0	—	5,400†	—	—	0.003	24	23			
0.5	0.5	2.6	5,000	0.04	1.5	0.004	22	63	0.5	0.5	—	7,000†	—	—	0.006	18	22				
	1.0	2.8	6,000	0.04	1.55	0.003	29	85		1.0	—	8,500†	—	—	0.003	23	23				
	2.0	2.9	6,200	0.04	1.5	0.003	27	100		2.0	—	9,650†	—	—	0.0015	26	23				
180	0.1	0.1	0.44	1,000	0.08	4.4	0.02	30	30	0.1	0.1	—	1,300†	—	—	0.03	35	19			
		0.25	0.5	1,200	0.08	4.4	0.015	52	41		0.25	—	1,700†	—	—	0.015	46	21			
		0.5	0.6	1,200	0.07	4.0	0.008	53	46		0.5	—	1,950†	—	—	0.007	50	22			
	0.25	0.25	1.18	1,900	0.05	2.7	0.01	39	55	0.25	0.25	—	2,950†	—	—	0.015	40	23			
		0.5	1.2	2,100	0.06	3.2	0.007	55	69		0.5	—	3,800†	—	—	0.007	50	24			
		1.0	1.5	2,200	0.05	3.0	0.003	53	83		1.0	—	4,300†	—	—	0.0035	57	24			
0.5	0.5	2.6	3,300	0.04	2.1	0.005	47	81	0.5	0.5	—	5,250†	—	—	0.007	44	24				
	1.0	2.8	3,500	0.04	2.0	0.003	55	115		1.0	—	6,600†	—	—	0.0035	54	25				
	2.0	3.0	3,500	0.04	2.2	0.002	53	116		2.0	—	7,650†	—	—	0.002	61	25				
300	0.1	0.1	0.5	950	0.09	4.6	0.025	60	36	0.1	0.1	—	1,150†	—	—	0.03	60	20			
		0.25	0.5	1,100	0.09	5.0	0.015	89	47		0.25	—	1,500†	—	—	0.015	83	22			
		0.5	0.6	900	0.08	4.8	0.009	86	54		0.5	—	1,750†	—	—	0.007	86	23			
	0.25	0.25	1.2	1,500	0.06	3.2	0.015	70	64	0.25	0.25	—	2,850†	—	—	0.015	75	23			
		0.5	1.2	1,600	0.06	3.5	0.008	100	79		0.5	—	3,400†	—	—	0.0055	87	24			
		1.0	1.5	1,800	0.08	4.0	0.004	95	100		1.0	—	4,000†	—	—	0.003	100	24			
0.5	0.5	2.7	2,400	0.05	2.5	0.006	80	96	0.5	0.5	—	4,850†	—	—	0.0055	76	23				
	1.0	2.9	2,500	0.05	2.3	0.003	120	150		1.0	—	6,100†	—	—	0.003	94	24				
	2.0	3.4	2,800	0.05	2.8	0.0025	90	145		2.0	—	7,150†	—	—	0.0015	104	24				

* The cathodes of the two units have a common terminal. † Values shown are for phase-inverter service.

6A06, 6AT6, 6Q7, 6Q7-G, 6Q7-GT, 6SL7-GT*, 6SZ7, 6T7-G, 12AT6, 12Q7-GT, 12SL7-GT* (see Circuit Diagram 1):

6AU6, 6SH7, 12AU6, 12SH7 (see Circuit Diagram 3):

E _{bb}	1)										E _{bb}	3)									
	R _p	R _g	R _{g2}	R _k	C _{g2}	C _k	C	E _o	V.G.	R _p		R _g	R _{g2}	R _k	C _{g2}	C _k	C	E _o	V.G.		
90	0.1	0.1	—	4,200	—	2.5	0.025	5.4	22†	0.1	0.1	0.07	1,800	0.11	9.0	0.021	25	52			
		0.22	—	4,600	—	2.2	0.014	7.5	27†		0.22	0.09	2,100	0.1	8.2	0.012	32	72			
		0.47	—	4,800	—	2.0	0.0065	9.1	30†		0.47	0.096	2,100	0.1	8.0	0.0065	37	88			
	0.22	0.22	—	7,000	—	1.5	0.013	7.3	30†	0.22	0.22	0.25	3,100	0.08	6.2	0.009	25	72			
		0.47	—	7,800	—	1.3	0.007	10	34†		0.47	0.26	3,200	0.078	5.8	0.0055	32	99			
		1.0	—	8,100	—	1.1	0.0035	12	37†		1.0	0.35	3,700	0.085	5.1	0.003	34	125			
0.47	0.47	—	12,000	—	0.83	0.006	10	36†	0.47	0.47	0.75	6,300	0.042	3.4	0.0035	27	102				
	1.0	—	14,000	—	0.7	0.0035	14	39†		1.0	0.75	6,500	0.042	3.3	0.0027	32	126				
	2.2	—	15,000	—	0.6	0.002	16	41†		2.2	0.8	6,700	0.04	3.2	0.0018	36	152				
180	0.1	0.1	—	1,900	—	3.6	0.027	19	30†	0.1	0.1	0.12	800	0.15	14.1	0.021	57	74			
		0.22	—	2,200	—	3.1	0.014	25	35		0.22	0.15	900	0.126	14.0	0.012	82	116			
		0.47	—	2,500	—	2.8	0.0065	32	37		0.47	0.19	1,000	0.1	12.5	0.006	81	141			
	0.22	0.22	—	3,400	—	2.2	0.014	24	38	0.22	0.22	0.38	1,500	0.09	9.6	0.009	59	130			
		0.47	—	4,100	—	1.7	0.0085	34	42		0.47	0.43	1,700	0.08	8.7	0.005	67	171			
		1.0	—	4,600	—	1.5	0.0035	38	44		1.0	0.6	1,900	0.066	8.1	0.003	71	200			
0.47	0.47	—	6,600	—	1.1	0.0085	29	44	0.47	0.47	0.9	3,100	0.06	5.7	0.0045	54	172				
	1.0	—	8,100	—	0.9	0.0035	38	46		1.0	1.0	3,400	0.05	5.4	0.0028	65	232				
	2.2	—	9,100	—	0.8	0.002	43	47		2.2	1.2	3,600	0.04	5.6	0.0019	72	272				
300	0.1	0.1	—	1,500	—	4.4	0.027	40	34	0.1	0.1	0.2	500	0.13	18.0	0.019	76	109			
		0.22	—	1,800	—	3.6	0.014	54	38		0.22	0.24	600	0.11	16.4	0.011	103	145			
		0.47	—	2,100	—	3.0	0.0065	63	41		0.47	0.26	700	0.11	15.3	0.006	129	168			
	0.22	0.22	—	2,600	—	2.5	0.013	51	42	0.22	0.22	0.42	1,000	0.1	12.4	0.009	92	164			
		0.47	—	3,200	—	1.9	0.0065	65	46		0.47	0.5	1,100	0.098	11.0	0.007	108	230			
		1.0	—	3,700	—	1.6	0.0035	77	48		1.0	0.55	1,000	0.09	11.0	0.003	122	262			
0.47	0.47	—	5,200	—	1.2	0.006	61	48	0.47	0.47	1.0	1,800	0.075	8.0	0.0045	94	248				
	1.0	—	6,300	—	1.0	0.0035	74	50		1.0	1.1	1,900	0.065	7.6	0.0028	105	318				
	2.2	—	7,200	—	0.9	0.002	85	51		2.2	1.2	2,100	0.06	7.3	0.0018	122	371				

* One triode unit. The cathodes of the two units have separate terminals. † At 2 volts (rms) output. ‡ At 3 volts (rms) output. ¶ At 4 volts (rms) output.

RC-coupled Amplifier Charts. (Continued)

6BF6, 6R7, 6R7-GT, 6SR7, 6ST7, 12SR7 (see Circuit Diagram 1):

6C4, 12AU7* (see Circuit Diagram 1):

E_{bb}	R_p	R_o	R_{g2}	R_k	C_{g2}	C_k	C	E_o	V.G.	E_{bb}	R_p	R_o	R_{g2}	R_k	C_{g2}	C_k	C	E_o	V.G.
90	0.047	0.047	—	2,200	—	2.5	0.083	14	9	90	0.047	0.047	—	1,600	—	3.2	0.061	9	10†
		0.1	—	2,800	—	2.0	0.033	18	10			0.1	—	1,800	—	2.5	0.033	11	11‡
		0.22	—	3,200	—	1.7	0.015	20	10			0.22	—	2,000	—	2.0	0.015	14	11
	0.1	0.1	—	4,100	—	1.4	0.032	13	10		0.1	0.1	—	3,000	—	1.6	0.032	10	11‡
		0.22	—	5,400	—	1.0	0.013	20	11			0.22	—	3,800	—	1.1	0.015	15	11
		0.47	—	6,400	—	0.9	0.007	24	11			0.47	—	4,500	—	1.0	0.007	18	11
0.22	0.22	—	8,500	—	0.67	0.015	18	11	0.22	0.22	—	6,800	—	0.7	0.015	14	11		
	0.47	—	12,000	—	0.5	0.0065	23	11		0.47	—	9,500	—	0.5	0.0065	20	11		
	1.0	—	14,000	—	0.43	0.0035	27	11		1.0	—	11,500	—	0.43	0.0035	24	11		
180	0.047	0.047	—	2,000	—	2.9	0.062	32	10	180	0.047	0.047	—	920	—	3.9	0.062	20	11
		0.1	—	2,500	—	2.2	0.033	42	10			0.1	—	1,200	—	2.9	0.037	26	12
		0.22	—	3,000	—	1.9	0.016	47	11			0.22	—	1,400	—	2.5	0.016	29	12
	0.1	0.1	—	3,800	—	1.5	0.033	36	11		0.1	0.1	—	2,000	—	1.9	0.032	24	12
		0.22	—	5,100	—	1.1	0.015	47	11			0.22	—	2,800	—	1.4	0.016	33	12
		0.47	—	6,200	—	0.9	0.007	55	12			0.47	—	3,600	—	1.1	0.007	40	12
0.22	0.22	—	8,000	—	0.73	0.015	41	12	0.22	0.22	—	5,300	—	0.8	0.015	31	12		
	0.47	—	11,000	—	0.5	0.007	54	12		0.47	—	8,300	—	0.56	0.007	44	12		
	1.0	—	13,000	—	0.4	0.0035	60	12		1.0	—	10,000	—	0.48	0.0035	54	12		
300	0.047	0.047	—	1,800	—	3.0	0.063	58	10	300	0.047	0.047	—	870	—	4.1	0.065	38	12
		0.1	—	2,400	—	2.4	0.033	74	11			0.1	—	1,200	—	3.0	0.034	52	12
		0.22	—	2,900	—	2.0	0.016	85	11			0.22	—	1,500	—	2.4	0.016	68	12
	0.1	0.1	—	3,600	—	1.6	0.033	65	12		0.1	0.1	—	1,900	—	1.9	0.032	44	12
		0.22	—	5,000	—	1.2	0.015	85	12			0.22	—	3,000	—	1.3	0.016	68	12
		0.47	—	6,200	—	0.95	0.007	96	12			0.47	—	4,000	—	1.1	0.007	80	12
0.22	0.22	—	7,800	—	0.73	0.015	74	12	0.22	0.22	—	5,300	—	0.9	0.015	57	12		
	0.47	—	11,000	—	0.5	0.007	95	12		0.47	—	8,800	—	0.52	0.007	82	12		
	1.0	—	13,000	—	0.43	0.0035	106	12		1.0	—	11,000	—	0.46	0.0035	92	12		

* One triode unit. The cathodes of the two units have separate terminals. † At 3 volts (rms) output. ‡ At 4 volts (rms) output.

6C5, 6C5-GT, 6C6*, 6J7*, 6J7-G*, 6J7-GT*, 6W7-G*, 12J7-GT*, 57* (see Circuit Diagram 1):

6C8-G† (see Circuit Diagram 1):

E_{bb}	R_p	R_o	R_{g2}	R_k	C_{g2}	C_k	C	E_o	V.G.	E_{bb}	R_p	R_o	R_{g2}	R_k	C_{g2}	C_k	C	E_o	V.G.
90	0.05	0.05	—	2,800	—	2.0	0.05	14	9	90	0.1	0.1	—	3,040	—	2.34	0.028	13	18
		0.1	—	3,400	—	1.62	0.025	17	9			0.25	—	3,700	—	1.48	0.0115	17	20
		0.25	—	3,800	—	1.3	0.01	20	10			0.5	—	4,520	—	1.29	0.006	19	21
	0.1	0.1	—	4,800	—	1.12	0.025	16	10		0.25	0.25	—	6,770	—	0.95	0.011	15	21
		0.25	—	6,400	—	0.84	0.01	22	11			0.5	—	7,870	—	0.81	0.0065	19	23
		0.5	—	7,500	—	0.66	0.005	23	12			1.0	—	8,830	—	0.69	0.0035	21	23
0.25	0.25	—	11,400	—	0.52	0.01	18	12	0.5	0.5	—	12,400	—	0.51	0.006	16	22		
	0.5	—	14,500	—	0.4	0.006	23	12		1.0	—	15,000	—	0.43	0.0035	20	24		
	1.0	—	17,300	—	0.33	0.004	26	13		2.0	—	16,500	—	0.38	0.0015	25	24		
180	0.05	0.05	—	2,200	—	2.2	0.055	34	10	180	0.1	0.1	—	2,420	—	2.34	0.028	30	20
		0.1	—	2,700	—	2.1	0.03	45	11			0.25	—	3,080	—	1.84	0.012	40	22
		0.25	—	3,100	—	1.85	0.015	54	11			0.5	—	3,560	—	1.6	0.0065	45	23
	0.1	0.1	—	3,900	—	1.7	0.035	41	12		0.25	0.25	—	5,170	—	1.25	0.012	35	24
		0.25	—	5,300	—	1.25	0.015	54	12			0.5	—	6,560	—	0.95	0.007	45	25
		0.5	—	6,200	—	1.2	0.008	55	13			1.0	—	7,550	—	0.85	0.0035	50	26
0.25	0.25	—	9,500	—	0.74	0.015	44	13	0.5	0.5	—	9,840	—	0.66	0.007	38	25		
	0.5	—	12,300	—	0.55	0.008	52	13		1.0	—	12,500	—	0.5	0.004	44	26		
	1.0	—	14,700	—	0.47	0.004	59	13		2.0	—	15,600	—	0.44	0.0015	51	26		
300	0.05	0.05	—	2,100	—	3.16	0.075	57	11	300	0.1	0.1	—	2,120	—	3.93	0.037	55	22
		0.1	—	2,600	—	2.3	0.04	70	11			0.25	—	2,840	—	2.01	0.013	73	23
		0.25	—	3,100	—	2.2	0.015	83	12			0.5	—	3,250	—	1.79	0.007	80	25
	0.1	0.1	—	3,800	—	1.7	0.035	65	12		0.25	0.25	—	4,750	—	1.29	0.013	64	25
		0.25	—	5,300	—	1.3	0.015	84	13			0.5	—	6,100	—	0.98	0.0065	80	26
		0.5	—	6,000	—	1.17	0.008	88	13			1.0	—	7,100	—	0.77	0.004	90	27
0.25	0.25	—	9,600	—	0.9	0.015	73	13	0.5	0.5	—	9,000	—	0.67	0.007	67	27		
	0.5	—	12,300	—	0.59	0.008	85	14		1.0	—	11,500	—	0.48	0.004	83	27		
	1.0	—	14,000	—	0.37	0.003	97	14		2.0	—	14,500	—	0.37	0.002	96	28		

* As triode. † One triode unit. The cathodes of the two units have separate terminals.

RC-coupled Amplifier Charts. (Continued)

6F8-G*, 6J5, 6J5-GT, 6SN7-GT*, 12J5-GT, A₅ Pentodes, 6C6, 6J7, 6J7-G, 6J7-GT, 6W7-G, 12SN7-GT* (see Circuit Diagram 1); 12J7-GT, 57 (see Circuit Diagram 3):

E_{bb}	R_p	R_g	R_{g2}	R_k	C_{g2}	C_k	C	E_o	V.G.	E_{bb}	R_p	R_g	R_{g2}	R_k	C_{g2}	C_k	C	E_o	V.G.
90	0.05	0.05	—	1,650	—	2.80	0.06	11	11	90	0.1	0.1	0.37	1,200	0.05	5.2	0.02	17	41
		0.1	—	2,070	—	2.66	0.029	14	12			0.25	0.44	1,100	0.05	5.3	0.01	22	55
		0.25	—	2,380	—	1.95	0.012	17	13			0.5	0.44	1,300	0.05	4.8	0.006	33	66
	0.1	0.1	—	3,470	—	1.85	0.035	12	13		0.25	0.25	1.1	2,400	0.03	3.7	0.008	23	70
		0.25	—	3,940	—	1.29	0.012	17	13			0.5	1.18	2,600	0.03	3.2	0.005	32	85
		0.5	—	4,420	—	1.0	0.007	19	13			1.0	1.4	3,600	0.025	2.5	0.003	33	92
0.25	0.25	—	7,860	—	0.73	0.0135	14	13	0.5	0.5	2.18	4,700	0.02	2.3	0.005	28	93		
	0.5	—	9,760	—	0.55	0.007	18	13		1.0	2.0	5,500	0.05	2.0	0.0025	29	120		
	1.0	—	10,690	—	0.47	0.004	20	13		2.0	2.7	5,500	0.02	2.0	0.0015	27	140		
180	0.05	0.05	—	1,190	—	3.27	0.06	24	13	180	0.1	0.1	0.44	1,000	0.05	6.5	0.02	42	51
		0.1	—	1,490	—	2.86	0.032	30	13			0.25	0.5	750	0.05	6.7	0.01	52	69
		0.25	—	1,740	—	2.06	0.0115	36	13			0.5	0.5	800	0.05	6.7	0.006	59	83
	0.1	0.1	—	2,330	—	2.19	0.038	26	14		0.25	0.25	1.1	1,200	0.04	5.2	0.008	41	93
		0.25	—	2,830	—	1.35	0.012	34	14			0.5	1.18	1,600	0.04	4.3	0.005	60	118
		0.5	—	3,230	—	1.15	0.006	38	14			1.0	1.4	2,000	0.04	3.8	0.0035	60	140
0.25	0.25	—	5,560	—	0.81	0.013	28	14	0.5	0.5	2.45	2,600	0.03	3.2	0.005	45	135		
	0.5	—	7,000	—	0.62	0.007	36	14		1.0	2.9	3,100	0.025	2.5	0.0025	56	165		
	1.0	—	8,110	—	0.5	0.004	40	14		2.0	2.7	3,500	0.02	2.8	0.0015	60	165		
300	0.05	0.05	—	1,020	—	3.56	0.06	41	13	300	0.1	0.1	0.44	500	0.07	8.5	0.02	55	61
		0.1	—	1,270	—	2.98	0.034	51	14			0.25	0.5	450	0.07	8.3	0.01	81	82
		0.25	—	1,500	—	2.15	0.012	60	14			0.5	0.53	600	0.06	8.0	0.006	96	94
	0.1	0.1	—	1,900	—	2.31	0.035	43	14		0.25	0.25	1.18	1,100	0.04	5.5	0.008	81	104
		0.25	—	2,440	—	1.42	0.0125	56	14			0.5	1.18	1,200	0.04	5.4	0.005	104	140
		0.5	—	2,700	—	1.2	0.0065	64	14			1.0	1.45	1,300	0.05	5.8	0.005	110	185
0.25	0.25	—	4,590	—	0.87	0.013	46	14	0.5	0.5	2.45	1,700	0.04	4.2	0.005	75	161		
	0.5	—	5,770	—	0.64	0.0075	57	14		1.0	2.9	2,200	0.04	4.1	0.003	97	200		
	1.0	—	6,950	—	0.54	0.004	64	14		2.0	2.95	2,300	0.04	4.0	0.0025	100	230		

* One triode unit. The cathodes of the two units have separate terminals.

6L5-G (see Circuit Diagram 1):

6S7, 6S7-G (see Circuit Diagram 3):

E_{bb}	R_p	R_g	R_{g2}	R_k	C_{g2}	C_k	C	E_o	V.G.	E_{bb}	R_p	R_g	R_{g2}	R_k	C_{g2}	C_k	C	E_o	V.G.
90	0.05	0.05	—	2,120	—	2.3	0.05	14	9.3	90	0.1	0.1	0.59	870	0.065	5.1	0.018	16	33
		0.1	—	2,500	—	1.86	0.03	18	10			0.25	0.65	900	0.061	5.0	0.01	21	47
		0.25	—	2,900	—	1.65	0.014	21	11			0.5	0.7	910	0.057	4.58	0.007	23	54
	0.1	0.1	—	3,510	—	1.36	0.03	16	11		0.25	0.25	1.5	1,440	0.044	3.38	0.007	14	56
		0.25	—	4,620	—	1.08	0.015	22	12			0.5	1.6	1,520	0.044	3.23	0.0053	18	66
		0.5	—	5,200	—	1.0	0.0085	23	12			1.0	1.7	1,560	0.043	3.22	0.004	19	77
0.25	0.25	—	8,050	—	0.61	0.0125	18	12	0.5	0.5	3.2	2,620	0.029	2.04	0.004	12	70		
	0.5	—	10,300	—	0.49	0.0085	22	12		1.0	3.5	2,800	0.03	1.95	0.0026	15	84		
	1.0	—	12,100	—	0.42	0.0055	24	12		2.0	3.7	3,000	0.031	1.92	0.0024	16	94		
180	0.05	0.05	—	1,810	—	2.9	0.06	32	10	180	0.1	0.1	0.58	530	0.073	7.2	0.017	33	47
		0.1	—	2,240	—	2.2	0.03	41	11			0.25	0.68	540	0.07	6.9	0.01	43	66
		0.25	—	2,660	—	1.8	0.014	46	12			0.5	0.71	540	0.065	6.6	0.0063	48	75
	0.1	0.1	—	3,180	—	1.46	0.03	36	12		0.25	0.25	1.6	850	0.05	4.6	0.0071	33	79
		0.25	—	4,200	—	1.1	0.0145	46	12			0.5	1.8	890	0.044	4.7	0.005	40	104
		0.5	—	4,790	—	1.0	0.009	50	12			1.0	1.9	950	0.046	4.4	0.0037	44	118
0.25	0.25	—	7,190	—	0.7	0.014	38	12	0.5	0.5	3.3	1,410	0.041	3.5	0.0041	30	109		
	0.5	—	9,290	—	0.54	0.009	46	12		1.0	3.6	1,520	0.037	3.0	0.003	38	134		
	1.0	—	10,950	—	0.46	0.0055	52	13		2.0	3.8	1,600	0.031	2.9	0.0024	42	147		
300	0.05	0.05	—	1,740	—	2.91	0.06	56	11	300	0.1	0.1	0.59	430	0.007	8.5	0.0167	57	57
		0.1	—	2,160	—	2.18	0.032	68	12			0.25	0.67	440	0.071	8.0	0.01	75	78
		0.25	—	2,600	—	1.82	0.015	79	12			0.5	0.71	440	0.071	8.0	0.0066	82	89
	0.1	0.1	—	3,070	—	1.64	0.032	60	12		0.25	0.25	1.7	620	0.058	6.0	0.0071	54	98
		0.25	—	4,140	—	1.1	0.014	79	13			0.5	1.95	650	0.057	5.8	0.005	66	122
		0.5	—	4,700	—	0.81	0.0075	89	13			1.0	2.1	700	0.055	5.2	0.0036	76	136
0.25	0.25	—	6,900	—	0.57	0.013	64	13	0.5	0.5	3.6	1,000	0.04	4.1	0.0037	52	136		
	0.5	—	9,100	—	0.46	0.0075	80	13		1.0	3.9	1,080	0.041	3.9	0.0029	66	162		
	1.0	—	10,750	—	0.4	0.005	88	13		2.0	4.1	1,120	0.043	3.8	0.0023	73	174		

* At 4 volts (rms) output.

RC-coupled Amplifier Charts. (Continued)

6SC7*, 12SC7* (see Circuit Diagram 4):

6F5, 6F5-GT, 6SF5, 6SF5-GT, 12F5-GT, 12SF5 (see Circuit Diagram 1):

E_{bb}	R_p	R_g	R_{g1}	R_k	C_{g1}	C_k	C	E_o	V.G.	E_{bb}	R_p	R_g	R_{g1}	R_k	C_{g1}	C_k	C	E_o	V.G.
90	0.1	0.1	—	1,850§	—	—	0.028	4.1	13†	90	0.1	0.1	—	4,400	—	2.5	0.02	4	28‡
		0.25	—	1,960§	—	—	0.012	5.9	23‡			0.25	—	4,800	—	2.1	0.01	5	34‡
		0.5	—	2,050§	—	—	0.0065	8.9	25¶			0.5	—	5,000	—	1.8	0.005	6	35¶
	0.25	0.25	—	3,400§	—	—	0.011	6.2	26¶		0.25	0.25	—	8,000	—	1.33	0.01	6	39¶
		0.5	—	5,750§	—	—	0.006	8.6	30			0.5	—	8,800	—	1.18	0.005	7	43¶
		1.0	—	3,900§	—	—	0.003	10	33			1.0	—	9,000	—	0.9	0.003	10	44
0.5	0.5	—	5,500§	—	—	0.005	7.4	31	0.5	0.5	—	12,200	—	0.76	0.005	8	43		
	1.0	—	6,300§	—	—	0.003	10	33		1.0	—	13,500	—	0.67	0.003	10	46		
	2.0	—	7,450§	—	—	0.0015	12	36		2.0	—	14,700	—	0.58	0.0015	12	48		
180	0.1	0.1	—	960§	—	—	0.031	17	25	180	0.1	0.1	—	1,800	—	4.4	0.025	16	37
		0.25	—	1,070§	—	—	0.012	24	29			0.25	—	2,000	—	3.3	0.015	23	44
		0.5	—	1,220§	—	—	0.0065	27	33			0.5	—	2,200	—	2.9	0.006	25	46
	0.25	0.25	—	1,850§	—	—	0.011	21	35		0.25	0.25	—	3,500	—	2.3	0.01	21	48
		0.5	—	2,150§	—	—	0.006	28	39			0.5	—	4,100	—	1.8	0.006	26	53
		1.0	—	2,400§	—	—	0.003	32	41			1.0	—	4,500	—	1.7	0.004	32	57
0.5	0.5	—	3,050§	—	—	0.006	24	40	0.5	0.5	—	6,100	—	1.3	0.006	24	53		
	1.0	—	3,420§	—	—	0.003	32	43		1.0	—	6,900	—	0.9	0.003	33	63		
	2.0	—	3,890§	—	—	0.002	36	45		2.0	—	7,700	—	0.83	0.0015	37	66		
300	0.1	0.1	—	750§	—	—	0.033	35	29	300	0.1	0.1	—	1,300	—	5.0	0.025	33	42
		0.25	—	930§	—	—	0.014	50	34			0.25	—	1,600	—	3.7	0.01	43	49
		0.25	—	1,040§	—	—	0.007	54	36			0.5	—	1,700	—	3.2	0.006	48	52
	0.25	0.25	—	1,400§	—	—	0.012	45	39		0.25	0.25	—	2,600	—	2.5	0.01	41	56
		0.5	—	1,680§	—	—	0.006	55	42			0.5	—	3,200	—	2.1	0.007	54	63
		1.0	—	1,840§	—	—	0.003	64	45			1.0	—	3,500	—	2.0	0.004	63	67
0.5	0.5	—	2,330§	—	—	0.006	50	45	0.5	0.5	—	4,500	—	1.5	0.006	50	65		
	1.0	—	2,980§	—	—	0.003	62	48		1.0	—	5,400	—	1.2	0.004	62	70		
	2.0	—	3,280§	—	—	0.002	72	49		2.0	—	6,100	—	0.93	0.002	70	70		

* The cathodes of the two units have a common terminal. † At 2 volts (rms) output. ‡ At 3 volts (rms) output. § Values are for phase-inverter service. ¶ At 4 volts (rms) output.

6SJ7, 12SJ7 (see Circuit Diagram 3):

6SJ7, 6SJ7-GT, 12SJ7, 12SJ7-GT (see Circuit Diagram 3):

E_{bb}	R_p	R_g	R_{g1}	R_k	C_{g1}	C_k	C	E_o	V.G.	E_{bb}	R_p	R_g	R_{g1}	R_k	C_{g1}	C_k	C	E_o	V.G.
90	0.1	0.1	0.26	1,500	0.11	4.8	0.02	21	21	90	0.1	0.1	0.29	820	0.09	8.8	0.02	18	41
		0.22	0.3	1,600	0.1	4.4	0.012	26	29			0.25	0.29	880	0.085	7.4	0.016	23	68
		0.47	0.35	1,900	0.09	4.2	0.006	28	37			0.5	0.31	1,000	0.075	6.6	0.007	28	70
	0.22	0.22	0.64	2,400	0.09	3.4	0.009	21	33		0.25	0.25	0.69	1,680	0.06	5.0	0.012	16	75
		0.47	0.7	2,500	0.09	3.2	0.0055	26	40			0.5	0.92	1,700	0.045	4.5	0.005	18	93
		1.0	0.84	2,600	0.084	3.0	0.0035	29	52			1.0	0.82	1,800	0.04	4.0	0.003	22	104
0.47	0.47	1.5	4,200	0.06	2.1	0.0045	21	50	0.5	0.5	1.5	3,600	0.045	2.4	0.003	18	91		
	1.0	1.6	4,400	0.06	1.9	0.003	26	59		1.0	1.7	3,800	0.03	2.4	0.002	22	119		
	2.2	1.7	4,800	0.058	1.6	0.002	29	64		2.0	1.9	4,050	0.028	2.85	0.0015	24	139		
180	0.1	0.1	0.33	1,000	0.13	6.7	0.02	32	33	180	0.1	0.1	0.29	760	0.10	9.1	0.019	49	55
		0.22	0.5	1,200	0.12	5.8	0.011	37	45			0.25	0.31	800	0.09	8.0	0.015	60	82
		0.47	0.6	1,300	0.11	5.5	0.006	43	52			0.5	0.37	860	0.09	7.8	0.007	62	91
	0.22	0.22	0.76	1,700	0.11	4.5	0.0095	37	47		0.25	0.25	0.83	1,050	0.06	6.8	0.001	38	109
		0.47	0.9	1,700	0.1	4.5	0.0055	44	68			0.5	0.94	1,090	0.06	6.6	0.004	47	131
		1.0	1.0	1,800	0.1	4.2	0.003	47	82			1.0	0.94	1,100	0.07	6.1	0.003	54	161
0.47	0.47	1.8	3,300	0.09	2.9	0.0045	38	70	0.5	0.5	1.85	2,000	0.05	4.0	0.003	37	151		
	1.0	2.0	3,800	0.08	2.4	0.003	50	85		1.0	2.2	2,180	0.04	3.8	0.002	44	192		
	2.2	2.1	4,000	0.07	2.3	0.002	57	98		2.0	2.4	2,410	0.035	3.6	0.0015	54	208		
300	0.1	0.1	0.32	750	0.19	8.0	0.021	62	39	300	0.1	0.1	0.35	500	0.10	11.6	0.019	72	67
		0.22	0.36	850	0.18	7.7	0.012	80	46			0.25	0.37	530	0.09	10.9	0.016	96	98
		0.47	0.37	900	0.18	7.7	0.006	93	57			0.5	0.47	590	0.09	9.9	0.007	101	104
	0.22	0.22	0.8	1,150	0.13	6.0	0.01	63	62		0.25	0.25	0.89	850	0.07	8.5	0.011	79	139
		0.47	0.94	1,300	0.12	6.7	0.0055	78	88			0.5	1.10	860	0.06	7.4	0.004	88	187
		1.0	0.98	1,500	0.11	6.0	0.0035	99	97			1.0	1.18	910	0.06	6.9	0.003	98	185
0.47	0.47	1.7	2,300	0.1	3.5	0.0045	71	82	0.5	0.5	2.0	1,300	0.06	6.0	0.004	64	200		
	1.0	1.9	2,500	0.1	3.5	0.003	89	109		1.0	2.2	1,410	0.05	5.8	0.002	79	238		
	2.2	2.0	2,800	0.09	3.1	0.002	105	125		2.0	2.5	1,530	0.04	5.2	0.0015	89	263		

RC-coupled Amplifier Charts. (Continued)

6Z7-G* (see Circuit Diagram 4):

55, 85 (see Circuit Diagram 1):

E_{bb}	R_p	R_g	R_{g2}	R_k	C_{g2}	C_k	C	E_o	V.G.	E_{bb}	R_p	R_g	R_{g2}	R_k	C_{g2}	C_k	C	E_o	V.G.	
90	0.1	0.1	—	1,480†	—	2.65	0.025	8	21†	90	0.05	0.1	—	3,800	—	1.4	0.06	16	4.5	
		0.25	—	1,760†	—	2.02	0.0115	11	25			0.25	—	4,600	—	1.1	0.03	19	4.9	
		0.5	—	1,930†	—	1.7	0.0065	14	26			0.25	—	5,400	—	0.86	0.015	23	5.1	
	0.25	0.25	—	3,000†	—	1.36	0.01	12	28		0.1	0.25	0.1	—	6,820	—	0.7	0.04	17	5.1
		0.5	—	3,390†	—	1.1	0.006	15	30				0.25	—	9,000	—	0.55	0.015	22	5.4
		1.0	—	3,670†	—	0.8	0.0035	18	33				0.5	—	10,300	—	0.5	0.007	25	5.5
0.5	0.5	—	5,300†	—	0.65	0.0055	14	31	0.25	1.0	0.25	—	15,100	—	0.31	0.015	18	5.3		
	1.0	—	6,050†	—	0.61	0.003	18	33			0.5	—	20,500	—	0.25	0.007	23	5.5		
	2.0	—	6,700†	—	0.45	0.0015	20	35			1.0	—	24,400	—	0.2	0.004	26	5.6		
180	0.1	0.1	—	930†	—	3.4	0.028	18	26	180	0.05	0.05	—	3,200	—	1.8	0.06	33	4.9	
		0.25	—	1,100†	—	2.6	0.0115	28	31			0.1	—	4,100	—	1.6	0.045	44	5.2	
		0.5	—	1,210†	—	2.32	0.007	33	32			0.25	—	5,000	—	1.2	0.02	49	5.3	
	0.25	0.25	—	1,820†	—	1.71	0.012	28	35		0.1	0.25	0.1	—	6,200	—	0.9	0.04	37	5.3
		0.5	—	2,110†	—	1.38	0.007	34	38				0.25	—	8,700	—	0.7	0.015	47	5.5
		1.0	—	2,400†	—	1.1	0.0035	41	39				0.5	—	10,000	—	0.57	0.008	50	5.5
0.5	0.5	—	3,240†	—	0.9	0.006	32	39	0.25	0.1	0.25	—	14,500	—	0.43	0.015	40	5.6		
	1.0	—	3,890†	—	0.703	0.0035	38	40			0.5	—	20,000	—	0.29	0.008	48	5.7		
	2.0	—	4,360†	—	0.553	0.002	44	41			1.0	—	24,000	—	0.24	0.004	53	5.7		
300	0.1	0.1	—	670†	—	3.81	0.028	38	31	300	0.05	0.05	—	3,200	—	1.9	0.08	50	5.2	
		0.25	—	950†	—	2.63	0.012	52	34			0.1	—	4,100	—	1.5	0.045	74	5.5	
		0.5	—	1,050†	—	2.34	0.007	60	36			0.25	—	5,100	—	1.2	0.015	85	5.6	
	0.25	0.25	—	1,430†	—	1.87	0.012	50	38		0.1	0.25	0.1	—	5,900	—	0.8	0.03	64	5.5
		0.5	—	1,680†	—	1.46	0.006	59	40				0.25	—	8,300	—	0.54	0.015	82	5.7
		1.0	—	1,930†	—	1.19	0.0035	66	43				0.5	—	9,600	—	0.43	0.006	88	5.8
0.5	0.5	—	2,540†	—	0.97	0.006	55	42	0.25	0.1	0.25	—	14,300	—	0.3	0.01	71	5.7		
	1.0	—	3,110†	—	0.72	0.0035	70	44			0.5	—	19,400	—	0.22	0.006	84	5.7		
	2.0	—	3,560†	—	0.56	0.002	75	45			1.0	—	23,600	—	0.2	0.003	94	5.8		

* The cathodes of the two units have a common terminal. † At 4 volts (rms) output. ‡ Values are for phase-inverter service.

56, 76 (see Circuit Diagram 1):

79* (see Circuit Diagram 4):

E_{bb}	R_p	R_g	R_{g2}	R_k	C_{g2}	C_k	C	E_o	V.G.	E_{bb}	R_p	R_g	R_{g2}	R_k	C_{g2}	C_k	C	E_o	V.G.	
90	0.05	0.05	—	2,500	—	2.0	0.06	16	7.0	90	0.1	0.1	—	2,050¶	—	—	0.04	5.8	23†	
		0.1	—	3,200	—	1.6	0.03	21	7.7			0.25	—	2,200¶	—	—	0.015	8.4	29†	
		0.25	—	3,800	—	1.25	0.015	23	8.1			0.5	—	2,350¶	—	—	0.009	9.5	29	
	0.1	0.1	—	4,500	—	1.05	0.03	19	8.1		0.25	1.0	0.25	—	4,000¶	—	—	0.015	7.1	31†
		0.25	—	6,500	—	0.82	0.015	23	8.9				0.5	—	4,250¶	—	—	0.006	9.7	33
		0.5	—	7,500	—	0.68	0.007	25	9.3				1.0	—	4,650¶	—	—	0.004	12	35
0.25	0.25	—	11,100	—	0.48	0.015	21	9.4	0.5	2.0	0.5	—	6,150¶	—	—	0.006	8.8	34		
	0.5	—	15,100	—	0.36	0.007	24	9.7			1.0	—	6,850¶	—	—	0.004	12	33		
	1.0	—	18,300	—	0.32	0.0035	28	9.8			2.0	—	7,500¶	—	—	0.002	15	40		
180	0.05	0.05	—	2,400	—	2.5	0.06	36	7.7	180	0.1	0.1	—	1,050¶	—	—	0.04	21	27	
		0.1	—	3,000	—	1.9	0.035	48	8.2			0.25	—	1,250¶	—	—	0.02	27	31	
		0.25	—	3,700	—	1.65	0.015	55	9.0			0.5	—	1,350¶	—	—	0.009	31	34	
	0.1	0.1	—	4,500	—	1.45	0.035	45	9.3		0.25	1.0	0.25	—	2,050¶	—	—	0.02	26	37
		0.25	—	6,500	—	0.97	0.015	55	9.5				0.5	—	2,450¶	—	—	0.01	34	41
		0.5	—	7,600	—	0.8	0.008	57	9.8				1.0	—	2,750¶	—	—	0.005	40	42
0.25	0.25	—	10,700	—	0.6	0.015	49	9.7	0.5	2.0	0.5	—	3,450¶	—	—	0.009	30	42		
	0.5	—	14,700	—	0.45	0.007	59	10			1.0	—	4,100¶	—	—	0.0035	39	44		
	1.0	—	17,700	—	0.4	0.0045	64	10			2.0	—	4,650¶	—	—	0.002	44	45		
300	0.05	0.05	—	2,400	—	2.8	0.08	65	8.3	300	0.1	0.1	—	800¶	—	—	0.025	40	29	
		0.1	—	3,100	—	2.2	0.045	80	8.9			0.25	—	1,000¶	—	—	0.01	57	34	
		0.25	—	3,800	—	1.8	0.02	95	9.4			0.5	—	1,100¶	—	—	0.006	60	36	
	0.1	0.1	—	4,500	—	1.6	0.04	74	9.5		0.25	1.0	0.25	—	1,650¶	—	—	0.01	56	39
		0.25	—	6,400	—	1.2	0.02	95	10				0.5	—	2,050¶	—	—	0.0055	66	42
		0.5	—	7,500	—	0.98	0.009	104	10				1.0	—	2,350¶	—	—	0.003	77	43
0.25	0.25	—	11,100	—	0.69	0.02	82	10	0.5	2.0	0.5	—	2,850¶	—	—	0.0055	61	44		
	0.5	—	15,200	—	0.5	0.009	96	10			1.0	—	3,600¶	—	—	0.003	75	46		
	1.0	—	18,300	—	0.4	0.005	108	10			2.0	—	4,450¶	—	—	0.0015	82	46		

* The cathodes of the two units have a common terminal. † At 3 volts (rms) output. ‡ At 4 volts (rms) output. ¶ Values are for phase-inverter service.

RC-coupled Amplifier Charts. (Continued)

12AX7* (see Circuit Diagram 1):

E_{bb}	R_p	R_g	R_{g2}	R_k	C_{g2}	C_k	C	E_o	V.G.
90	0.1	0.1	—	4,400	—	2.7	0.023	5	29†
		0.22	—	4,700	—	2.4	0.013	6	35†
		0.47	—	4,800	—	2.3	0.007	8	41†
	0.22	0.22	—	7,000	—	1.6	0.001	6	39†
		0.47	—	7,400	—	1.4	0.006	9	45‡
		1.0	—	7,600	—	1.3	0.003	11	48‡
	0.47	0.47	—	12,000	—	0.9	0.006	9	48‡
		1.0	—	13,000	—	0.8	0.003	11	52‡
		2.2	—	14,000	—	0.7	0.002	13	55‡
180	0.1	0.1	—	1,800	—	4.0	0.025	18	40
		0.22	—	2,000	—	3.5	0.013	25	47
		0.47	—	2,200	—	3.1	0.006	32	52
	0.22	0.22	—	3,000	—	2.4	0.012	24	53
		0.47	—	3,500	—	2.1	0.006	34	59
		1.0	—	3,900	—	1.8	0.003	39	63
	0.47	0.47	—	5,800	—	1.3	0.006	30	62
		1.0	—	6,700	—	1.1	0.003	39	66
		2.2	—	7,400	—	1.0	0.002	45	68
300	0.1	0.1	—	1,300	—	4.6	0.027	43	45
		0.22	—	1,500	—	4.0	0.013	57	52
		0.47	—	1,700	—	3.6	0.006	66	57
	0.22	0.22	—	2,200	—	3.0	0.013	54	59
		0.47	—	2,800	—	2.3	0.006	69	65
		1.0	—	3,100	—	2.1	0.003	79	68
	0.47	0.47	—	4,300	—	1.6	0.006	62	69
		1.0	—	5,200	—	1.3	0.003	77	73
		2.2	—	5,900	—	1.1	0.002	92	75

* One triode unit. † At 2 volts (rms) output. ‡ At 3 volts (rms) output. ¶ At 4 volts (rms) output.

CHAPTER 9

RADIO-FREQUENCY AMPLIFIERS

BY R. S. GLASGOW¹

1. Types. Radio-frequency amplifiers are designed to operate above the audible range of frequencies. The mode of operation may be either class A, class B, or class C. Class A operation is employed when voltage output is the primary consideration and is the type employed in receiving sets and many types of electronic laboratory equipment. This mode is characterized by freedom from distortion and relatively low power output. Class B and class C amplifiers are used when power output is the important factor. Their principal use is in radio transmitters; class B being used when the impressed signal voltage has undergone amplitude modulation, and class C when amplitude modulation is absent in the signal to be amplified. The plate efficiency is much higher than with class A operation, approaching 90 per cent in large class C amplifiers.

Radio-frequency amplifiers may be required to amplify a comparatively narrow band of frequencies, which is the usual requirement in radio reception, or an extremely

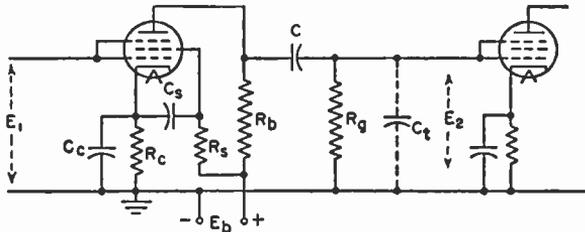


FIG. 1. Resistance-coupled amplifier.

wide band extending from several megacycles down to the lower values of audio frequency, as in video-frequency (v-f) amplifiers. In the former case the tuning may be either fixed or continuously adjustable over the tuning range of the receiver. Provision is sometimes made to vary the width of the band to be amplified.

2. Resistance-coupled Voltage Amplifiers. Where uniform gain is to be secured over a wide range of frequencies, pentodes employing resistance coupling between stages are generally used. Higher amplification per stage can be obtained than with triodes and with a much smaller grid input capacitance C_g . The latter item, augmented by the distributed capacitance of the wiring plus the output capacitance of previous tube, is the factor that causes the inevitable reduction in gain at the higher frequencies.

A simplified circuit is shown in Fig. 1, omitting the various compensating schemes usually employed to secure better performance at the high or low ends of the frequency range with regard to gain or phase shift between input and output voltages. The capacitance C_t represents the combined shunting effect of the various capacitances mentioned in the preceding paragraph. The variation of amplification with frequency of this circuit is shown in

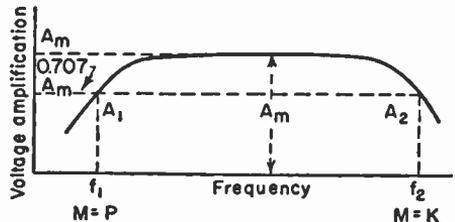


FIG. 2. Variation of voltage amplification in a resistance-coupled amplifier.

¹ U. S. Postgraduate School, Annapolis, Md.

Fig. 2 with frequency plotted on a logarithmic scale. The useful frequency range is usually taken as lying between f_1 and f_2 . These two values are the so-called "half-power points" where the voltage amplification A has fallen to 70.7 per cent (-3 db) below the mid-frequency value A_m where the gain is substantially constant with frequency. They correspond to the cutoff frequencies of a band-pass filter, although in the case of the amplifier there is not such an abrupt change in gain as with a well-designed filter.

The reduction in gain at low frequencies is caused by the increased reactance of the coupling capacitor C in comparison to R_g . In this lower range of frequencies the effect of C_i is negligible. Assuming that the by-pass capacitors C_c and C_s are of sufficient size so that their reactances are small compared to R_c and R_s , respectively, the voltage amplification in the lower frequency range is given by

$$A_L = \frac{E_2}{E_1} = \frac{\mu R_b R_g}{(r_p R_b + r_p R_g + R_b R_g) - jX_c(r_p + R_b)} = \frac{\mu R_b R_g}{|M - jP|} \tag{1}$$

where $X_c = 2\pi fC$
 μ = tube amplification factor
 r_p = tube plate resistance
 $M = r_p R_b + r_p R_g + R_b R_g$
 $P = X_c(r_p + R_b)$

Where $M = P$, the magnitude of the denominator in Eq. (1) is $|M - jP| = \sqrt{2M}$, as shown in Fig. 3. Hence, when $M = P$ the voltage amplification at f_1 is

$$A_1 = \frac{\mu R_b R_g}{\sqrt{2} M} = 0.707 A_m \tag{2}$$

The value of the lower cutoff frequency is

$$2\pi f_1 = \frac{A_m(r_p + R_b)}{\mu R_b R_g C} = \frac{A_m}{g_m R_b R_g C} + \frac{A_m}{\mu R_g C} \tag{3}$$

where g_m is the transconductance of the tube and is equal to μ/r_p . The phase shift at f_1 will be 45 deg leading. At other frequencies in the low range, the phase shift ϕ is given by

$$\phi = \tan^{-1}(P/M) \tag{4}$$

From Eq. (3) it will be seen that a low value of f_1 requires a small value of A_m and a large value for C .

At the middle range of frequencies the reactance of C can be neglected in comparison to R_g , and the load in the plate circuit becomes effectively R_b in parallel with R_g . The reactance of C_i is still high enough to have no appreciable shunting effect across the equivalent resistance of R_b and R_g in parallel. The voltage amplification in this region will be substantially constant and is given by

$$A_m = \frac{\mu R_b R_g}{r_p R_b + r_p R_g + R_b R_g} = \frac{\mu R_b R_g}{M} \tag{5}$$

Since r_p under the usual operating conditions of a pentode is apt to be in the vicinity of a megohm, and so large in comparison to the equivalent load resistance in the plate circuit, Eq. (5) can be written as

$$A_m = g_m \frac{R_b R_g}{R_b + R_g} \tag{6}$$

to a close degree of approximation. The phase shift is substantially zero over the frequency range for which Eq. (5) is valid, and will extend from about one octave above f_1 to one octave below f_2 .

At frequencies above the middle range the reduction of the plate load impedance by

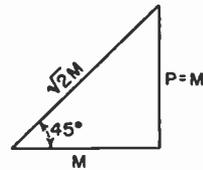


FIG. 3. Vector relations in Eq. (3) at frequency f_1 .

the increased shunting effect of C_i becomes greater, and the expression for the voltage amplification will be

$$A_k = \frac{E_2}{E_1} = \frac{\mu R_b R_g}{(r_p R_b + r_p R_g + R_b R_g) + j\omega C_i r_p R_b R_g} = \frac{\mu R_b R_g}{|M + jK|} \quad (7)$$

where $K = \omega C_i r_p R_b R_g$

When $M = K$, the magnitude of the denominator in Eq. (7) is again $\sqrt{2} M$, and A_2 in Fig. 2 will be given by Eq. (2), as before. The frequency f_2 at the half-power point is

$$f_2 = \frac{g_m}{2\pi C_i A_m} \quad (8)$$

Consequently, to obtain a high value of f_2 the transconductance of the tube used should be as large as possible, while A_m must be kept small. The minimum value of C_i that can ordinarily be secured (including the circuit connections) is about $25 \mu\mu\text{f}$ with pentodes now available. The most suitable of these have a value of g_m of about 9,000 μmhos , so that a voltage amplification A_m of about 5 to 10 per stage is all that can be employed if f_2 is to be several megacycles. In this case R_b will be only a few thousand ohms. The figure of merit for a tube to be used in a cascade amplifier, based upon the maximum frequency that can be amplified, is seen in Eq. (8) to be the ratio of the transconductance to the sum of the input and output capacitances.

Since r_p in pentodes is large compared to the usual load impedance in the plate circuit, the expression for the gain in Eq. (7) is closely approximated by

$$A_k = \frac{g_m R_b R_g}{(R_b + R_g) \sqrt{1 + (f^2/f_2^2)}} \quad (9)$$

At frequencies higher than the middle range, the phase shift is lagging and is given by

$$\phi = \tan^{-1}(K/M) \quad (10)$$

The angle will be 45 deg at the upper cutoff frequency f_2 .

3. Grid Input Capacitance. One item of the capacitance C_i in Fig. 1 is the tube input capacitance C_g . In the case of a pentode it is given by the equation

$$C_g = C_{gf} + C_{gs} + C_{gp}(1 + A) \quad (11)$$

where C_{gf} = grid-cathode capacitance

C_{gs} = grid-screen capacitance

C_{gp} = grid-plate capacitance

A = voltage amplification of the tube

The last term of Eq. (11) is the smallest of the three because of the very small value of C_{gp} in pentodes. In tuned r-f amplifiers having high gain, this term may amount to 1 or 2 $\mu\mu\text{f}$. Variations in the amplification A , as are produced by automatic gain control (which varies g_m by changing either the grid bias or the voltage applied to the screen grid) will accordingly vary C_g by a small amount. The detuning effect thus caused must be guarded against in i-f amplifiers, particularly if the band of frequencies transmitted is relatively narrow or if the intermediate frequency is high. This is ordinarily taken care of by seeing that the total tuning capacitance employed is large compared to the variations in C_g . Where the band of frequencies to be amplified is relatively wide, as in the i-f amplifier of television receivers, the variations in C_g can usually be ignored.

4. Compensated Resistance-coupled Amplifiers. In Fig. 2 it is seen that the gain falls off as f_2 is approached, as given by Eq. (9). This gradual reduction in gain is frequently undesirable and a more sharply defined cutoff frequency f_0 would be preferable, particularly for design purposes. One method of securing a higher cutoff, and also a more constant gain characteristic up to f_0 , is to place a small inductance L in series with the plate-circuit load resistance R , as shown in Fig. 4. This value of load

resistance is $R = R_b R_o / (R_b + R_o)$, but as R_o is usually large compared to R_b , the actual value of R_b may be substituted for R in the equations that follow.

At low frequencies the reactance of L is small, and the load impedance is essentially equal to R . But as the frequency increases, the impedance of the branch $R + j\omega L$ becomes progressively greater, thus tending to offset the shunting effect of C_t .

The plate-load impedance is given by

$$Z_L = \frac{R + j\omega[L - C_t(R^2 + \omega^2 L^2)]}{\omega^2 C_t^2 R^2 + (\omega^2 L C_t - 1)^2} \quad (12)$$

and it may be shown that, by making the reactance of the coil L equal to half the load resistance R at the limiting frequency f_o , Z_L will be substantially constant in magnitude and equal to R . The gain will be approximately $g_m R$ up to a frequency of f_o . The required relations are

$$R = \frac{1}{2\pi f_o C_t} = 4\pi f_o L \quad (13)$$

Substituting Eq. (13) in Eq. (12), the load impedance becomes

$$Z_L = \frac{R \left[1 - j \left(\frac{f^2}{4f_o^2} + \frac{f}{2f_o} \right) \right]}{\left(\frac{f}{f_o} \right)^2 + \left(\frac{f^2}{2f_o^2} - 1 \right)^2} \quad (14)$$

The accurate expression for the voltage amplification per stage is then

$$A = \frac{g_m R \sqrt{1 + \left(\frac{f^2}{4f_o^2} + \frac{f}{2f_o} \right)^2}}{\left(\frac{f}{f_o} \right)^2 + \left(\frac{f^2}{2f_o^2} - 1 \right)^2} \quad (15)$$

Using the circuit constants discussed in Sec. 2, following Eq. (8), L will be in the vicinity of $25 \mu\text{h}$. Additional compensation methods are described in the reference below,¹ and in Chap. 19 on Television.

The relationship between the circuit constants given in Eq. (13) produces a small rise in the gain as f_o is approached. Where several stages of amplification are employed, as is usually the case, the cumulative effect of this "hump" in the over-all gain characteristic may be too great if identical stages are employed. Consequently, the usual practice is to vary the constants slightly in the individual stages so that the rise in the gain characteristic does not occur at the same frequency in each stage. This staggering of the individual characteristics can be brought about by choosing a slightly different set of values for the constants in Eq. (13) so as to secure the over-all characteristic desired. If the gain per stage is expressed in decibels where

$$\text{db} = 20 \log_{10} A \quad (16)$$

the over-all gain at the particular frequency is merely the sum of the decibel gains of the individual stages at that frequency.

5. Tuned Impedance-coupled Amplifier. The simplest form of tuned amplifier which employs a parallel resonant circuit as the plate load impedance is shown in Fig. 5. In receiving circuits, this form of coupling is often used in i-f amplifiers of superheterodynes that require a high value of intermediate frequency, as in u-h-f receiving sets and radars. In the latter the values of intermediate frequency may

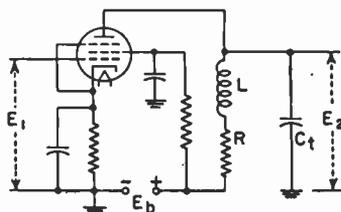


FIG. 4. Compensated resistance-coupled amplifier.

¹ Seeley and Kimball, Analysis and Design of Video Amplifiers, RCA Rev., pp. 290-308, January, 1939.

extend up to a mid-band value of 60 Mc. Here, the requirement for the transmission of a relatively wide band of frequencies (2 to 10 Mc) usually necessitates staggered tuning of the individual stages, which may be six or more in number. The complications involved in securing the proper tuning alignment require that the tuning adjustment of each stage be made as simple as possible, which favors the single adjustment of the circuit of Fig. 5 as against the more complicated adjustments of band-pass filter circuits.

The circuit of Fig. 5, with various modifications in the details, is commonly used as the coupling means between amplifier stages in radio transmitting circuits. C is not usually variable, except in low-power transmitters, and tuning is accomplished by varying the position of taps on the coil L .

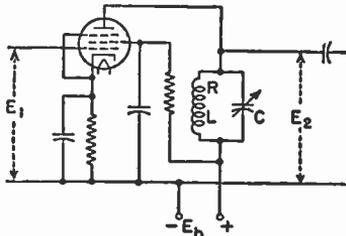


FIG. 5. Tuned impedance-coupled amplifier.

A vernier adjustment of inductance is provided in the form of a heavy copper or aluminum disk which acts as a short-circuited secondary of a single turn and which can be rotated within the coil. By rotating the plane of this disk, a fine adjustment of inductance is obtained for tuning purposes. The position of the taps should be chosen so that the plane of the disk is displaced nearly 90 deg, at resonance, from that of the coil, so as to minimize the I^2R loss in the disk.

In receiving circuits, particularly where space may be at a premium, a fixed capacitor may be used for C , tuning being accomplished by adjusting the position of a suitable cylindrical core of molded iron dust which is arranged so that it can be moved in and out of the coil. Improvements in the quality of these iron-dust cores in recent years enables them to be used to advantage at much higher frequencies. Values of Q ($= \omega L/R$) in the vicinity of 100 can be obtained in the broadcast range of frequencies, using a coil diameter of about $3/8$ in. and a length of about 1 in. This method, known as *permeability tuning*, has been used in connection with push-button tuning, particularly in automotive receiving sets.

The voltage amplification of the circuit in Fig. 5 at any frequency is given by

$$A = \frac{E_2}{E_1} = g_m Z_L \tag{17}$$

where Z_L is given in vector form by Eq. (12). Transforming Eq. (12) into its scalar magnitude, Eq. (17) becomes

$$A = g_m \sqrt{\frac{R^2 + \omega^2 L^2}{\omega^2 C^2 R^2 + (\omega^2 LC - 1)^2}} \tag{18}$$

$$= g_m \sqrt{\frac{\omega^2 L^2 (1 + Q^2)}{\omega^4 L^2 C^2 + Q^2 (\omega^2 LC - 1)^2}}$$

At resonance the voltage amplification will be

$$A = g_m \frac{R^2 + \omega^2 L^2}{R} = g_m \omega L \frac{1 + Q^2}{Q} \tag{19}$$

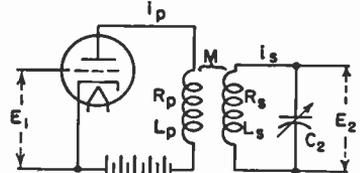


FIG. 6. Tuned transformer-coupled amplifier.

6. Tuned Transformer-coupled Amplifiers. A typical tuned transformer-coupled amplifier is shown in Fig. 6. Receiving sets of the t-r-f type are seldom used today as the great majority now use the superheterodyne circuit. However, most of the better receivers of the latter type employ a stage of t.r.f. ahead of the first detector. In its original form, which is still used to some extent in some of the short-wave bands of an all-wave receiver, the primary coil L_p in Fig. 6 was of lower inductance than the secondary. More uniform selectivity may be obtained over the tuning range by using a primary inductance large enough to give the primary circuit—in conjunction with its own distributed capacitance plus the output capacitance of the tube—a resonant

frequency which lies somewhat below the l-f tuning limit of the secondary. The secondary inductance for a frequency range of 1,600 to 550 kc is ordinarily from 200 to 250 μ h.

Since the resistance R_p and the reactance ωL_p of the primary can be neglected in comparison to the plate resistance r_p of the tube, the voltage E_2 across the secondary at any frequency is

$$E_2 = \frac{-jE_1\mu M}{C_2 \left[r_p R_s + \omega^2 M^2 + jr_p \left(\omega L_s - \frac{1}{\omega C_2} \right) \right]} \tag{20}$$

At resonance, where $\omega L_s = 1/\omega C_2$, the voltage amplification becomes

$$A = \frac{E_2}{E_1} = \frac{\mu\omega^2 M L_s}{r_p R_s + \omega^2 M^2} \tag{21}$$

If the mutual inductance M in Eq. (21) is adjusted to satisfy the condition

$$\omega M = \sqrt{r_p R_s} \tag{22}$$

the optimum value of voltage amplification will be obtained and Eq. (21) reduces to

$$A_{\text{opt}} = \frac{\mu\omega L_s}{2\sqrt{r_p R_s}} \tag{23}$$

which is the maximum amplification it is possible to obtain with a given tube and coil.

When M is adjusted to its optimum value, it will be noted that the figure of merit of the tube is $\mu/\sqrt{r_p}$. Therefore, if two tubes have equal values of transconductance, the one having the higher amplification factor will give the greater gain. Tetrodes and pentodes will accordingly produce a greater gain than a triode. With M less than optimum the gain becomes more nearly proportional to the transconductance of the tube. When optimum coupling is employed, the amplification is directly proportional to the ratio of the coil reactance to the square root of its resistance, instead of Q_s of the coil. With values of M considerably less than optimum, as when pentodes are used, the gain becomes more nearly proportional to the figure of merit Q_s of the coil.

The impedance looking into the primary coil in Fig. 6 is

$$Z_p' = R_p + j\omega L_p + \frac{\omega^2 M^2}{R_s + j \left(\omega L_s - \frac{1}{\omega C_2} \right)} \tag{24}$$

At resonance, with optimum coupling, $Z_p' = r_p$ of the tube. This condition differs from the resistance- and impedance-coupled amplifiers in that, in the latter two, optimum amplification is approached by making the impedance of the load very large compared to r_p .

If a pentode is used in the circuit of Fig. 6, the above equations are still applicable. Since these tubes have plate resistances r_p approaching a megohm in value and amplification factors varying from several hundred to several thousand, the coupling that can be used between primary and secondary without causing instability in the form of oscillations is far below the optimum value. The preceding equations can, therefore, be simplified. Since $r_p \gg \omega M$ in the case of a pentode, the expression for the secondary voltage in Eq. (20) becomes

$$E_2 = \frac{-jE_1 g_m M}{C_2 \left[R_s + j \left(\omega L_s - \frac{1}{\omega C_2} \right) \right]} \tag{25}$$

and the voltage amplification at resonance is

$$A = g_m Q_s \omega M \tag{26}$$

These tubes enable values of amplification per stage to be obtained which are much larger than can be obtained with triodes. With a given secondary coil the selectivity in the case of a pentode is better than with a triode; or for equal selectivities the pentode circuit can use a smaller and less expensive coil. These advantages, together with their freedom from oscillation without the use of neutralizing circuits, have caused ordinary triodes to be virtually abandoned in the field of r-f amplifiers for receiving circuits. Triodes are still used in the higher power amplifier stages of radio transmitters.

7. Combinations of Inductive and Capacitive Coupling. To secure better performance in tuned amplifiers without resorting to moving parts other than the tuning capacitors, combinations of inductive and capacitive coupling between stages have been used.¹ By a proper choice of circuit elements it is possible to make the effective coupling vary with the frequency in a predetermined manner. In this way the variation of gain with frequency can be given almost any desired characteristic.

Two examples of such circuits are shown in Fig. 7. In Fig. 7a the coil L_b has a large value of inductance; hence its distributed capacitance C_1 , augmented by C_{pf} of the

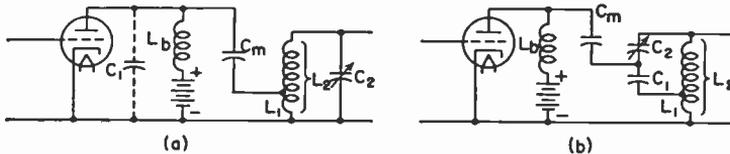


FIG. 7. Tuned amplifiers using combinations of inductive and capacitive coupling.

tube, resonates it to a frequency somewhat below the tuning range of the set. The output current of the tube divides between L_b and the path through C_m . At low frequencies a larger portion of the output current flows through this second path because of the high impedance offered by L_b as parallel resonance is approached in the latter. This causes the voltage induced in L_2 to remain more nearly constant over the tuning range.

The circuit of Fig. 7b accomplishes the same results in a somewhat different manner. L_b and C_m merely serve as choke coil and blocking capacitor of an amplifier using parallel feed. The amplified output current divides between C_1 and C_2 and then recombines to flow through the primary L_1 of the autotransformer. The capacitance of C_2 is increased as the signal frequency is lowered, which causes a progressive increase in the effective coupling. C_1 is about twenty times larger than the maximum value of C_2 , while L_1 includes about a turn or two of the coil L_2 .

Various types of filter circuits can also be used as a coupling means between stages. These refinements are usually impractical in receiving circuits owing to production difficulties in adjusting the circuit constants to the values required.

8. Cascade Amplifiers. If two or more identical stages of amplification are connected in cascade, the over-all voltage amplification is given by

$$A = A^n \quad (27)$$

where n = number of stages

A = amplification per stage

This expression presumes that the various stages do not react on each other, which is not always the case in practice owing to small unavoidable couplings between input and output circuits. If the various stages are not all identical, the over-all amplification will be the product of the individual values of A per stage. The response curve of a multistage amplifier composed of identical stages is readily obtained from the curve of an individual stage by raising its ordinates to the n th power, where n is the number of stages.

¹ Wheeler, H. A., and W. A. McDonald, Theory and Operation of Tuned Radiofrequency Coupling Systems, *Proc. IRE*, 19, 738, May, 1931.

When the amplification is expressed in decibels [Eq. (16)], it is only necessary to add the gains of the individual stages for the frequency in question to determine the overall gain.

The use of several stages of cascade t-r-f amplification enables both the selectivity and fidelity of the amplifier to be increased, provided the tuning of each stage is made broader as the number of stages is increased. This is illustrated in Fig. 8, both amplifier circuits being alike except for the values of mutual inductance between the primary and secondary of the t-r-f transformers. The necessity for broader tuning per stage in multistage amplifiers to avoid too great a sacrifice in fidelity permits the use of coils of rather compact dimensions wound with relatively small wire. The increased coil resistance thus produced will reduce the gain per stage, but this can be

offset if necessary by increasing the mutual inductance. At frequencies sufficiently remote from resonance where the gain per stage becomes less than unity, a cascade amplifier acts as an attenuator of the signal. An increase in the number of stages will, therefore, actually decrease the strength of interfering signals whose frequencies are above or below the band where the gain per stage is equal to or greater than one. All signals whose frequencies lie within this band will be strengthened by an increase in the number of stages. For this reason two types of selectivity may be recognized: the *adjacent-channel selectivity* and the *distant-channel selectivity*. It is, therefore, possible in a comparative test of two amplifiers of equal sensitivity to find that the first will produce less interference from interfering signal of, say, 30 kc away from resonance than the second; while for a signal of, say, 60 kc away there may be more interference present than in the second amplifier.

The attenuation of signals remote from the resonant frequency requires that the amplifier be well shielded to prevent short portions of the lead wires and circuits of the output stage from acting as antennas and picking up energy. Thus a few inches of exposed wire running to the grid of the detector tube might have a voltage induced in it from an interfering powerful local station which is much greater in magnitude than these same signals after passing through the amplifier.

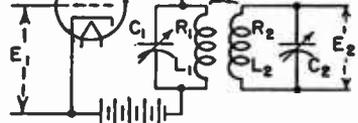


FIG. 9. Transformer-coupled amplifier with primary and secondary tuned.

curves is given in Fig. 10. Owing to the more uniform amplification obtained over a wider band of frequencies, these circuits are often referred to as band-pass filters. This form of circuit is commonly used in the i-f amplifier of superheterodynes.

When the primary and secondary are both tuned to the same frequency, the width of the transmitted band depends upon the magnitude of the coupling between them, and also upon the value of Q for the coils, but to a much less extent. A double-humped response curve results if M is greater than the critical value, and, as M is increased, the two peaks move farther apart and the hollow between them becomes deeper, particularly if the value of Q for the two coils is high.

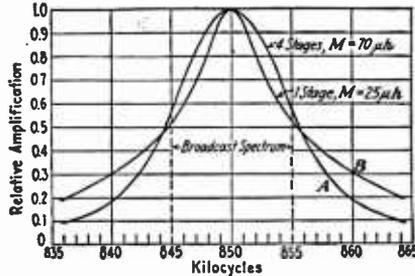


FIG. 8. Increase in selectivity with cascading.

9. Band-pass Filters. A rectangular response curve would be ideal for the r-f amplifier of a receiving set designed for entertainment purposes. The use of a pair of tuned circuits as a coupling means between stages results in a flatter response curve with steeper sides than can be obtained with a single tuned circuit. Such an arrangement is shown in Fig. 9, and the general appearance of the resultant response

In practice, both primary and secondary are tuned to the same frequency; consequently $\omega L_1 = 1/\omega C_1$ and $\omega L_2 = 1/\omega C_2$. If the resultant common resonant frequency is called f_0 , the selectivity characteristic can be determined, assuming a pentode to be used, from

$$E_2 = \frac{g_m E_1 M}{\omega C_1 C_2} \times \frac{1}{R_1 R_2 \left[1 - 4Q_1 Q_2 \left(\frac{f - f_0}{f_0} \right)^2 + j(Q_1 + Q_2) \frac{f - f_0}{f_0} \right] + \omega^2 M^2} \quad (28)$$

where f = frequency in question

Q_1 and Q_2 = values of $\omega L/R$ of primary and secondary circuits

At resonance when $f = f_0$, the voltage amplification will be

$$A = \frac{g_m M}{\omega C_1 C_2 (R_1 R_2 + \omega^2 M^2)} \quad (29)$$

A single-humped curve results when $\omega M = \sqrt{R_1 R_2}$, as shown in Fig. 10. In the case of this figure the value of Q for the two circuits is somewhat higher than would

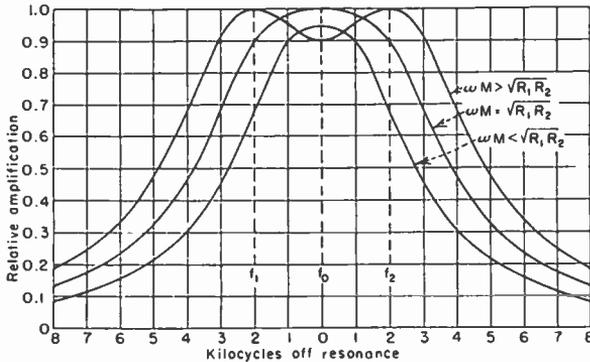


FIG. 10. Response curves of doubly tuned r-f amplifier.

be employed in the i-f amplifier of a receiving set designed for entertainment purposes. Although the selectivity would be excellent, the resulting attenuation of side bands in a two-stage amplifier using three such transformers would greatly impair the fidelity of reception. It is sometimes the practice to use transformers of slightly different characteristics in the several stages. Thus one or two of the transformers may have a more or less pronounced hollow at f_0 , while the other may have a single hump. In this way the amplification may be made fairly uniform throughout the band between f_1 and f_2 and then fall off sharply on either side. In television receivers where the i-f amplifier is called upon to transmit uniformly a band of frequencies about 4.5 Mc in width, a transformer consisting of three tuned coupled circuits may sometimes be required to secure the desired uniformity over the transmitted band. Staggered tuning of the various stages is also used to obtain a better over-all characteristic. The design problems involved in these applications are discussed in the references below.¹

With the trend toward higher fidelity in the better grade of broadcast receivers, it is highly desirable to have some adjustment control over the shape of the i-f response curve so as to be able to increase the fidelity on local reception when high selectivity to prevent interference is not required. But the broad response curve required would be unsatisfactory in many cases of distant reception where high selectivity might be

¹ Mountjoy, G., Television Signal-frequency Circuit Considerations, *RCA Rev.*, October, 1939, p. 204; and Simplified Television I-F Systems, *ibid.*, January, 1940, p. 299.

needed to avoid adjacent channel interference. One scheme of securing adjustable selectivity would be to vary the coupling between the primary and secondary coils by mechanical means, but practical difficulties prevent its use. Another method, illustrated in Fig. 11, is to have a small coil L_3 tightly coupled to L_1 . By rotating the switch arm to points 2 and 3, a progressive increase in M is secured between L_1 and the secondary circuit. This will result in a slight detuning of the secondary circuit as the response curve is widened, but this is of no serious consequence. Individual iron-dust cores in L_1 and L_2 enable close coupling to be had between L_1 and L_2 with a comparatively small number of turns in the latter.

10. Tuning Stability and Adjustments. The tuning alignment of the above circuits is often accomplished by using a fixed value of C and varying L by altering the position of the iron cores relative to the coils, rather than to employ an adjustable capacitor. The stability of some types of the latter over a period of time is apt to be unsatisfactory when subjected to a number of cyclic changes of temperature and humidity. This is particularly true of adjustable capacitors that use a flat leaf spring of metal as one plate, which is insulated from the other plate by means of a mica film. The capacitance is increased by means of a machine screw which squeezes the plates closer together.

A further advantage of permeability tuning is that a fixed capacitor may be used which has the proper negative temperature coefficient of capacitance, if necessary, and thereby compensate for the usual positive temperature coefficient of inductance possessed by most coils. Ceramic capacitors are available which have a wide variation of temperature coefficients, ranging from a small positive value to comparatively large negative values, depending upon the composition of their dielectrics. This property is very useful when it is necessary to preserve accurately the resonant alignment of tuned circuits over a considerable range of operating temperatures.

To utilize this property successfully it is essential that the values of the circuit elements vary in a cyclic fashion with temperature; *i.e.*, they return to their original values after each cycle of temperature variation.

Ceramic capacitors are sometimes used in parallel with other types so as to offset the positive temperature coefficients of the latter. However, cyclic behavior is difficult to secure in capacitors constructed by assembling alternate sheets of mica and metal foil in the form of a stack which is then clamped under pressure and impregnated, owing to changes in the pressure with temperature. Coating the opposite sides of the mica sheets with silver, applied in the form of a paste which is then baked, does away with this erratic behavior. These silvered mica capacitors are very stable

and have a small positive temperature coefficient. Variable air capacitors are apt to exhibit noncyclic performance unless proper precautions are observed in their design and construction.

11. Regeneration in Amplifiers. The three-electrode vacuum tube is not a perfect unilateral device but permits the amplified output energy to react upon

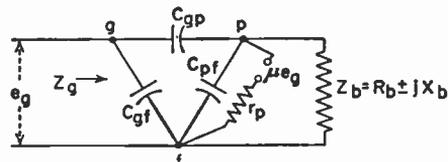


FIG. 12. Equivalent circuit of a triode.

the input circuit. The grid-to-plate capacitance of the tube serves to couple electrostatically the input and output circuits as shown in Fig. 12. If some of the output voltage is fed back into the input circuit so as to be in phase with e_g , the total, or regenerative amplification, may be expressed by

$$A_r = A \frac{\beta}{1 - A\beta} \tag{30}$$

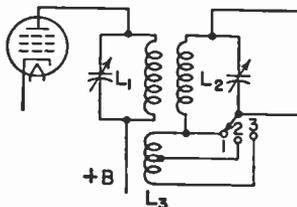


FIG. 11. Method for varying band width in i-f amplifier.

where β is the fraction of the output which is fed back into the input circuit and A is the gain of the amplifier if feedback were absent. If the quantity $A\beta$ is unity, the total amplification becomes infinite, and a continuous oscillation will result.

The oscillation of a single-stage amplifier can occur only if the plate circuit is sufficiently inductive. If the impedance in the plate circuit is pure resistance or a capacitive reactance, no oscillations can take place, although in the latter case anti-regenerative feedback may occur of sufficient magnitude greatly to reduce the resultant gain. The effect of feedback may be looked upon as being due to the input impedance Z_g of the grid-filament terminals of the tube. This impedance is of the form

$$Z_g = \pm r_g - j \frac{1}{\omega C_g} \tag{31}$$

When the plate circuit is inductive, the sign of r_g is negative, so that the tube is then capable of annulling part or all of the positive resistance of the associated circuit. In the latter event, oscillations occur. The effect of the various circuit elements of Fig. 12 on Z_g is given by

$$Z_g = \frac{C_{gp} + C_{pf} - j \frac{1}{\omega} \left(\frac{1}{R_b \pm jX_b} + \frac{1}{r_p} \right)}{\frac{\mu C_{gp}}{r_p} + (C_{gf} + C_{gp}) \left(\frac{1}{R_b \pm jX_b} + \frac{1}{r_p} \right) + j\omega(C_{gf}C_{gp} + C_{gp}C_{pf} + C_{pf}C_{gf})} \tag{32}$$

When Z_i is capacitive and has sufficient resistance associated with it, r_g is positive, and the tube may introduce rather large losses into the input circuit, even though the grid is biased sufficiently negative so that no conductive grid current flows.

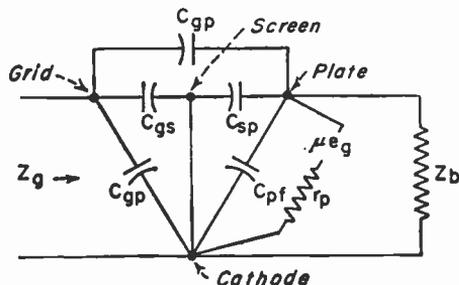


FIG. 13. Equivalent circuit of a tetrode or pentode.

Tetrodes and pentodes designed for r-f amplification are relatively immune from feedback through the grid-to-plate capacitance within the tube. The equivalent circuit, which applies to either of these types, is shown in Fig. 13. Comparing this circuit with that of the triode of Fig. 12, it is seen that the capacitance between the plate and the control grid is now broken up in effect into two capacitances in series, C_{sp} and C_{sg} , with their mid-point grounded to the cathode, so far as r-f potentials are concerned. The suppressor grid in a pentode, which is inserted between the plate and screen grid, is also at the cathode potential and augments the shielding action of the screen grid. The result is that the capacitance C_{gp} between control grid and plate is reduced to a value of less than 0.01 μmf in receiving tubes. Feedback of amplified output energy through the tube is thereby reduced to the point where stable operation can be obtained up to frequencies of several hundred megacycles.

These tubes may break into oscillation if too high a value of gain per stage is attempted, particularly as the operating frequency becomes higher. The reactance of C_{gp} diminishes with frequency and causes a progressive rise of the feedback factor β in Eq. (30). Consequently A must be reduced in order that $A\beta$ may remain less than unity. Capacitive coupling between external leads to the plate and control grid, or between components connected to these leads, must be kept as small as possible by shielding and by using care in the location of the wiring and circuit components. The use of staggered tuning, mentioned previously, where the input circuit of the tube is tuned to a slightly different frequency than the output circuit, is of considerable assistance in reducing the tendency to oscillate, particularly in a multistage amplifier. The improved stability, plus the more uniform amplification obtainable over a wider band of frequencies, is often of greater importance than the loss of over-all gain entailed

by this method. Troubles with instability increase somewhat in geometrical progression with the number of stages used. Individual leads from by-pass capacitors should be provided to the cathode of the tubes rather than to employ a common conductor for this purpose. The impedance of lead wires becomes of increasing consequence as the frequency becomes higher, as 1 in. of wire has roughly 1 ohm per Mc of reactance.

Indiscriminate circuit connections to ground at various locations on a metal chassis should also be avoided, as the paths taken by the r-f currents will overlap to some extent and cause a transfer of energy from a later stage in the amplifier to an earlier stage, thus causing regeneration. It must be borne in mind that a metal sheet is not an equipotential surface of zero impedance at radio frequencies. Another source of regeneration which must be guarded against in high-gain amplifiers is coupling caused by the use of a common plate-voltage source. Adequate decoupling circuits in the plate supply of each stage will prevent this cause of instability.

12. Grounded-grid Amplifiers. This type of circuit, also called a grid-separation circuit, enables stable amplifier operation to be secured with a triode up to the micro-

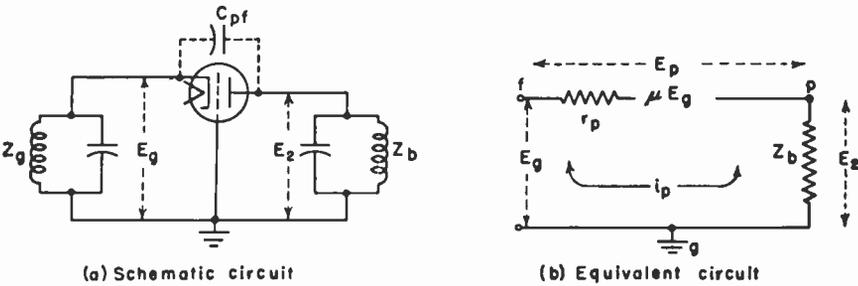


FIG. 14. Grounded-grid amplifier.

wave region without the use of neutralizing circuits. Specially designed triodes of the lighthouse or disk-seal type must be used in the upper portion of this frequency range. The basic circuit diagram is shown in Fig. 14a and the equivalent circuit in b. The signal voltage E_g is applied to the grid-filament terminals in the conventional manner, but the load circuit Z_b is connected between plate and grid, the latter electrode being grounded. The control grid thus serves as a screen between input and output circuits in somewhat the same manner as the screen grid of a tetrode. By using a grid construction of rather fine mesh and extending the structure so as to form a complete shield between plate and cathode, it is possible to reduce C_{pf} to values ranging from about $1 \mu\mu f$ in very large power tubes, to $0.05 \mu\mu f$ in small lighthouse tubes.

It will be observed from the equivalent circuit that the signal voltage E_g , the output terminals of the tube, and the load impedance Z_b are all in series. The amplifying action of the tube acts as a series booster to E_g , so that the voltage E_2 impressed across Z_b is $E_g + E_p$, where E_p is the a-c output voltage of the tube. The source of E_g accordingly supplies a portion of the total power to Z_b , even under class A operation, since the output current i_p flows through this source. Negative feedback is inherent in the grounded-grid circuit, the amount depending upon the internal impedance contained in the source of E_g . This means that β in Eq. (30) is a negative quantity.

The alternating plate current in Fig. 13 is given by

$$i_p = \frac{(\mu + 1)E_g}{r_p + Z_b} \tag{33}$$

and the voltage amplification is

$$A = \frac{E_2}{E_g} = \frac{(\mu + 1)Z_b}{r_p + Z_b} \tag{34}$$

The input impedance Z_g under class A operation differs from a conventional amplifier in that it is finite even at frequencies where the shunting effects of interelectrode

tube capacitances are negligible. Neglecting C_{gf} , it is given by

$$Z_g = \frac{E_g}{i_p} = \frac{Z_b + r_p}{\mu + 1} \quad (35)$$

Consequently, power equal to $E_g i_p$ must be supplied by the driver.

At higher frequencies where tube capacitance becomes a factor, the input impedance is lowered by the shunting effect of C_{gf} in parallel with Z_g . It should be noted that the input capacitance in the grounded-grid circuit is much lower than in the conventional triode or pentode amplifier, as it is not augmented by the additional terms of Eq. (11). The reduced values of input and output capacitance which result in this circuit, together with the constructional features of the lighthouse tube, enable the operating frequency to be extended up to several thousand megacycles. Extremely close electrode spacings of only a few mils between cathode and grid in this type of tube reduce the effect of electron transit time and are of further assistance in raising the upper frequency limit.

The operation of pentodes and neutralized triodes in the conventional amplifier circuits becomes increasingly difficult as the frequency is raised to the point where the inductances of lead wires becomes an important factor. It is then no longer possible

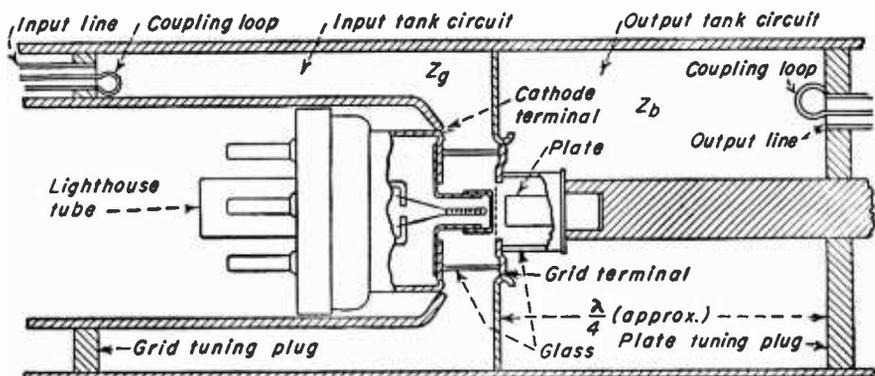


FIG. 15. Grounded-grid amplifier for u.h.f. using concentric lines as tuning elements.

to hold the screen and suppressor grids at r-f ground potential because of the inductive leads to these electrodes, and oscillation may occur. In a triode the leads to the external neutralizing capacitor and other circuit elements cause the frequency band over which neutralization can be secured to become progressively narrower as the frequency is increased, until the band finally vanishes altogether.

These difficulties with lead inductances and distributed capacitance at ultrahigh frequencies can be surmounted by the use of resonant lines as the tuned elements. Furthermore, if a concentric line is used, the circuit can be made self-shielding and energy losses by electromagnetic radiation can be avoided. Since a quarter-wavelength line short-circuited at the far end is electrically similar in its properties to a parallel resonant circuit, the equivalent of Fig. 14a is readily obtained by the construction illustrated in Fig. 15. The tube is incorporated into the structure as shown. Tuning of the input and output tank circuits is accomplished by varying the position of the adjustable tuning plugs. External coupling to these circuits is made by concentric lines terminated in small coupling loops as shown. With a type 2C40 lighthouse tube in a circuit of this type the maximum frequency that can be amplified is in the vicinity of 1,200 Mc.

The grounded-grid amplifier can also be operated class B or class C. More excitation power will be required from the driver than in conventional power amplifiers. The following power relations apply, assuming Z_b to be a pure resistance R_b :

$$\text{Power output} = (E_p + E_g)I_p = I_p^2 R_c \tag{36}$$

$$\text{Driver power} = E_g(I_p + I_g) \tag{37}$$

$$\text{Grid loss} = E_g I_g \tag{38}$$

where E_p and E_g are the fundamental components of the plate and grid voltages and I_p and I_g are the fundamental components of the plate and grid currents.

13. Cathode Followers. This name is commonly given to the cathode-coupled amplifier, the basic circuit of which is shown in Fig. 16. Either triodes or pentodes can be used, although in the latter case the screen and suppressor grids should be returned to the cathode instead of to ground. Since the output voltage E_2 is inserted into the input circuit in phase opposition to the signal voltage E_s , the circuit is completely degenerative and the feedback factor in Eq. (30) is $\beta = -1$. This results in a voltage amplification A which approaches, but cannot exceed, unity as a limiting factor. This disadvantage, which can usually be compensated by the use of adequate voltage gain preceding the cathode follower, is more than offset by its other advantages: (1) an adjustable value of output impedance that can be lowered to values of less than 50 ohms, (2) a reduced input capacitance, (3) reduced amplitude and frequency distortion over a very wide range of frequencies as a result of negative feedback, (4) freedom from instability when triodes are used.

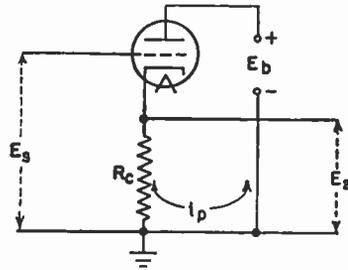


FIG. 16. Cathode-follower circuit.

The voltage amplification is given by

$$A = \frac{\mu R_c}{r_p + (\mu + 1)R_c} \tag{39}$$

In pentodes where μ is large compared to unity, Eq. (39) becomes

$$A = \frac{g_m R_c}{1 + g_m R_c} \tag{40}$$

where r_p = tube plate resistance

μ = tube amplification factor

$g_m = \mu/r_p$ = tube transconductance

Owing to negative feedback the equivalent plate resistance of the tube r_p' is

$$r_p' = \frac{r_p}{\mu + 1} \tag{41}$$

which is the theoretical value that the load resistance R should have to obtain maximum power output. If R_c is a coupling resistance across which a load such as a coaxial cable is to be connected, the output resistance presented to the cable terminals will be the product over the sum of r_p' and R_c and is

$$\text{Output impedance} = \frac{r_p R_c}{r_p + (\mu + 1)R_c} \tag{42}$$

In the case of pentodes, Eq. (42) becomes

$$\text{Output impedance} = \frac{R_c}{1 + g_m R_c} \tag{43}$$

The above relations enable a cathode follower to be conveniently matched to the load presented by a cable or other transmission line. The circuit thus functions in a manner similar to a transformer without the serious limitations that the latter would possess if operation over a wide range of frequencies were attempted.

The effective input capacitance C_g' as viewed from the source of E_s is reduced from

its actual value of C_g by the relation

$$C_g' = C_g(1 - A) \quad (44)$$

where A is given by Eq. (39) or (40). This reduces considerably the capacitive loading offered to the driver stage.

Cathode followers are used principally as an output stage.

14. Neutralizing Circuits. These circuits are used to prevent the oscillation of triode amplifiers caused by amplified output energy being fed into the input circuit through the grid-plate capacitance C_{gp} as discussed in Sec. 11. An additional function is to prevent reaction on the driver stage caused by tuning adjustments in the output of the driven stage. Neutralizing circuits of various types were quite widely used in receiving sets prior to the advent of pentode tubes. The principal use of these circuits at the present time is in connection with the large triodes employed in radio transmitters where suitable tetrodes and pentodes of the required characteristics or power ratings are not available.

The usual form of circuit isolated electrically the input and output terminals by converting them into two pairs of opposite points of a balanced a-c bridge. A type originated by C. W. Rice and widely used in power amplifiers is shown in Fig. 17, the sche-

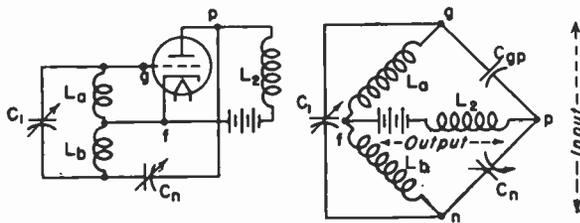


FIG. 17. Rice neutralized amplifier.

matic circuit being shown at the left and the electrical equivalent, omitting the tube electrodes, at the right. The filament terminal of the tube, instead of being connected to the lower end of the input circuit, is connected to an intermediate point which divides the inductance into two parts, L_a and L_b . The lower terminal n of the input circuit is connected to the plate through a small balancing capacitor C_n . The terminals g and n of the input circuit and f and p of the output circuit constitute two pairs of opposite points of a bridge. An inspection of the latter figure indicates that no voltage can exist across the input terminals gn due to a voltage between fp if the arms are balanced. Hence the energy which is fed back through C_{gp} is opposed in phase by that which flows through C_n . The conditions for a balance are

$$\frac{L_a}{L_b} = \frac{C_n}{C_{gp}} \quad (45)$$

This balance is not entirely independent of frequency as Eq. (45) would indicate unless the coupling between L_a and L_b is substantially unity. This is because L_a is shunted by the input capacitance of the tube. With certain arrangements a h-f parasitic oscillation may take place which will impair the performance of the amplifier at the frequencies for which it was designed. A small capacitance of about the size of C_n shunted across L_2 will often prevent such parasitics in receiving circuits. With perfect neutralization the input capacitance will be $C_{gf} + C_{gp} + C_n$ and will remain constant with adjustments in the plate load.

Another form of balancing circuit originated by L. A. Hazeltine and known as the *neutrodyne* is shown in Fig. 18. This type of circuit applies the same principle to the output circuit as the previous method did to the input. The conditions for balance are the same as Eq. (45). The coupling between L_a and L_s should again be approxi-

mately unity if the circuit is to remain balanced for a wide range of frequencies with a fixed adjustment of C_n , as L_a is shunted by the output impedance of the tube.

A circuit wherein all four of the bridge arms are capacitors is shown in Fig. 19. The grid-plate capacitance as well as the grid-filament capacitance of the tube is involved,

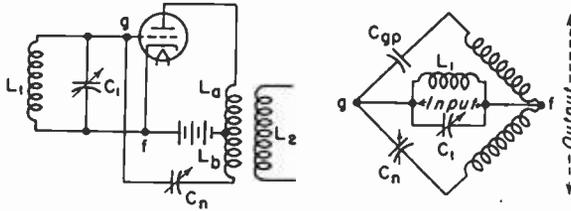


FIG. 18. Hazeltine neutralized amplifier.

these two capacitances serving as a pair of ratio arms. The conditions for a balance are

$$\frac{C_n}{C_a} = \frac{C_{gp}}{C_{gf}} \tag{46}$$

The value of C_a is usually about 100 $\mu\mu\text{f}$, which requires a value of C_n somewhat larger

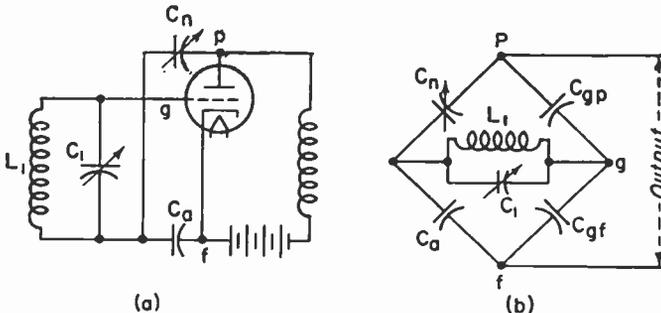


FIG. 19. Capacitance bridge neutralization of grid-plate capacitance.

in size than the neutralizing capacitors of the preceding circuits. To avoid the accumulation of a charge on the grid which may cause the tube to "block," C_a is usually shunted by a 250,000-ohm resistor. The distributed capacitance of a suitable choke coil whose natural frequency is below the frequency to be amplified can also be substituted for C_a .

Another form of circuit involving the principle of a mutual inductance bridge is illustrated in Fig. 20. The conditions for a balance are

$$\frac{M}{L_2} = \frac{C_{gp}}{C_{gf} + C_n} \tag{47}$$

Since C_n is in parallel with the grid-filament capacitance of the tube, it is possible to utilize C_{gf} in place of an actual neutralizing capacitor C_n and balance by proper adjustment of the mutual inductance between L_n and L_2 .

Push-pull circuits may be neutralized by the application of the circuit of Fig. 17, as in Fig. 21.

A different type of circuit where neutralization at only one frequency is desired, is shown in Fig. 22. The coil L_n is resonant with C_{gp} to form a parallel-resonant circuit

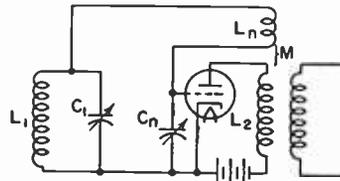


FIG. 20. Mutual inductance bridge circuit.

of high impedance between plate and grid and thereby exclude output energy from being fed back into the input circuit. The blocking capacitor C is required to prevent the d-c plate potential from being impressed on the grid.

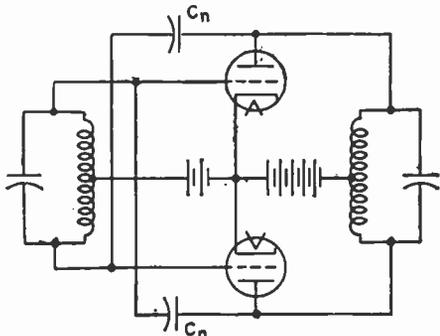


FIG. 21. Push-pull neutralizing circuit.

tube to be neutralized rather than to switch off the filament.

Figure 23 shows the last two stages of power amplification of a typical 1-kw broadcast transmitter. The first stage consists of two 75-watt screen-grid tubes in parallel which require no neutralization. The second stage is neutralized by means of C_n , which connects to the input tank circuit L_1C_1 at the point shown. The principle is the same as that of Fig. 17. The turns to which the various taps on L_1 are connected are indicated by the numbers. A 30-ohm resistance R_2 is connected in series with C_n to secure a more exact phase balance, since C_{gp} of the tube will have some losses associated with it and will therefore have a phase angle of less than 90 deg.

The neutralizing adjustment is made as follows: The switch S_1 is thrown to the top position inserting a low-range thermocouple Th_1 in the output tank circuit L_2C_2 . At the same time the galvanometer A_4 is connected to the thermocouple, and the plate circuit is opened by S_2 which is mechanically connected with S_1 . With excitation applied to the grid, C_n is then adjusted until A_4 reads zero. The switch

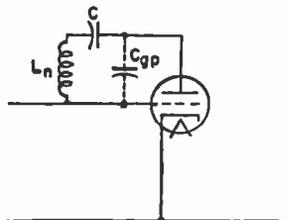


FIG. 22. Resonant neutralizing circuit.

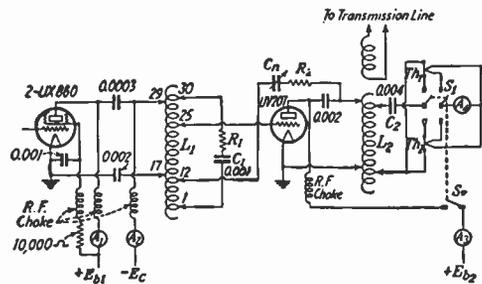
S_1 is then thrown to the lower position, closing the plate circuit and inserting a high-range thermocouple Th_2 in the tank circuit, and at the same time transferring A_4 .

16. Radio-frequency Power Amplifiers. The low output and plate efficiency of class A amplifiers preclude their use in transmitters, and class B or class C operation is employed.

Class B amplifiers are operated with a negative bias approximately equal to cutoff so that the plate current is almost zero when the alternating grid excitation is removed. With a sinusoidal voltage applied to the grid, the plate current consists of a series of half-sine waves, similar to the output of a half-wave rectifier. The load impedance is adjusted so as to obtain an approximately linear

FIG. 23. Broadcast transmitter power amplifier.

alternating grid excitation is removed. With a sinusoidal voltage applied to the grid, the plate current consists of a series of half-sine waves, similar to the output of a half-wave rectifier. The load impedance is adjusted so as to obtain an approximately linear



dynamic characteristic, as shown in Fig. 24. The grid swings positive on excitation peaks, causing grid current to flow. Class B amplifiers are used in radiotelephone transmitters following the modulated stage. The power output obtainable from a given tube is much greater than with class A operation and the plate efficiency is much higher, having a theoretical maximum value of 78.54 per cent. As with a-f power amplifiers, tubes operating as class B r-f amplifiers may also be operated in push-pull.

A class C amplifier is one in which high output and plate efficiency are the primary considerations. The grid is negatively biased to a point considerably beyond cutoff, as shown in Fig. 25, so that the plate current is zero with no grid excitation. The latter is quite large and is often sufficient to cause the plate current to reach saturation on positive swings. obtained with the larger tubes. the plate current to flow during

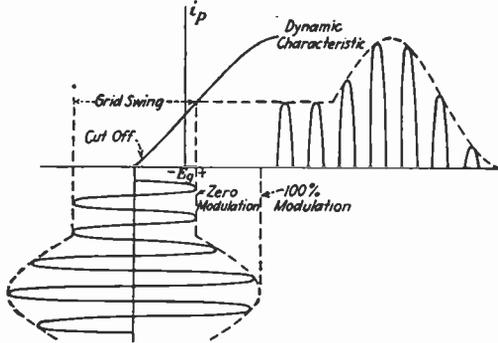


FIG. 24. Characteristics of class B amplification.

Plate efficiencies in the vicinity of 90 per cent may be These high efficiencies are made possible by allowing less than 180 deg of the cycle and only at a time when the plate potential is comparatively low. In radiotelegraph transmitters all stages are operated class C, while with radiotelephony only the modulated amplifiers and the stages preceding it are so operated.

The plate-current wave shapes in both cases are badly distorted, particularly with class C operation, and the output contains both odd and even harmonics. However, the tank circuit L_0C_0 in Fig. 26 is resonant to the fundamental to which it offers a high impedance of the nature of a pure resistance. The impedance offered to the plate-current harmonics diminishes rapidly with the order of the latter; hence the voltage drop E_0 across the tank circuit is very

nearly sinusoidal in shape. The instantaneous plate voltage e_p will be the algebraic difference between the plate-supply voltage E_b and the drop E_0 across the load.

Either triodes or screen-grid tetrodes may be used as power amplifiers. The latter have the advantage of not requiring neutralization. The screen-grid voltage in transmitting tubes is usually about 15 per cent of the plate-supply voltage, which is proportionately much lower than in receiving tubes. These tubes are difficult to construct in the larger sizes, and their operation is not too satisfactory, so that triodes must be used where large outputs are required.

17. Current and Voltage Relations. The instantaneous current and voltage relations for a class C amplifier are shown in Fig. 27. The potential e_p of the plate with respect to the filament is at a minimum during the time plate current is actually flowing. The power loss within the tube will be equal to the product of e_p and i_p averaged over a complete cycle. It is evident from Fig. 27 that this loss can be kept small by limiting the angle $2\theta_1$ during which time plate current actually flows. This will vary from 180 deg in

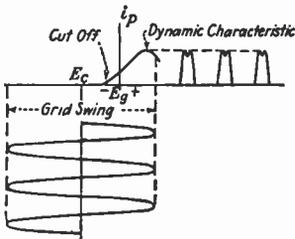


FIG. 25. Class C operation.

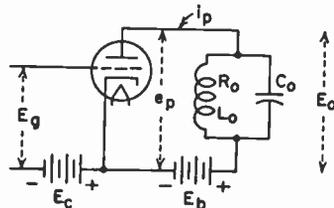


FIG. 26. Circuit of r-f power amplifier.

the case of a class B amplifier to perhaps as low as 60 deg for class C operation. It will also be noted that the grid-excitation voltage E_g is at its positive maximum when the plate voltage is a minimum. The minimum plate voltage should not be allowed to fall below the value of $e_g \text{ max}$ if excessive grid current is to be avoided. Ordinarily $e_g \text{ max}$ is limited to about 80 per cent $E_p \text{ min}$.

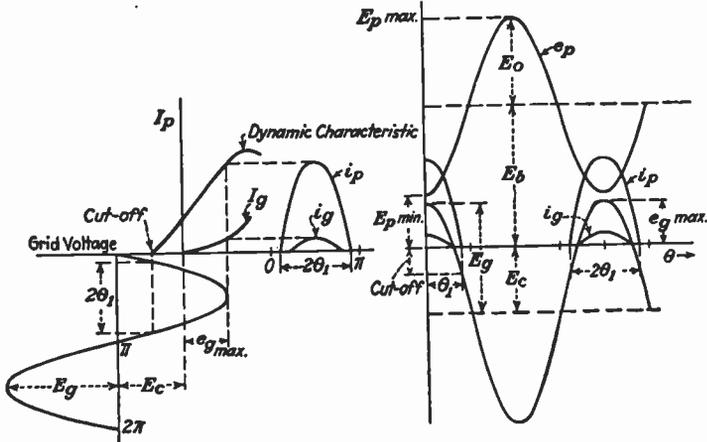


FIG. 27. Instantaneous values of current and voltage in class C amplifier.

18. Circuit Calculations. In the design of a power amplifier the given data will include the frequency, the type of tube to be used, and the plate-supply voltage. The minimum plate voltage and the maximum positive value of the grid voltage are then selected, also the angle θ_1 . The required grid-excitation voltage will be

$$E_g = \frac{E_b}{\mu} + \frac{1}{1 - \cos \theta_1} \left(\frac{E_p \text{ min} \cos \theta_1}{\mu} + e_g \text{ max} \right) \tag{48}$$

The required C bias will be

$$E_c = E_g - e_g \text{ max} \tag{49}$$

and the voltage across the tank circuit is given by

$$E_o = E_b - E_p \text{ min} \tag{50}$$

Corresponding pairs of plate and grid voltages can then be computed for increments of 5 to 10 deg over the time interval $2\theta_1$, during which plate current flows. Since the various current and voltage waves are symmetrical on either side of the vertical axis, it is only necessary to do this from zero to θ_1 . Table 1 is suitable for this purpose.

The values of plate and grid currents in lines 7 and 8 are obtained from the static characteristics of the tube for the computed pairs of instantaneous values of e_p and e_g in lines 4 and 6. The grid-current characteristic will also be necessary if the power required for grid excitation is to be determined.

The d-c component of plate current I_b will be the average value of i_p over a complete cycle and is given by

$$I_b = \frac{1}{18} \left(\frac{y_0}{2} + y_1 + y_2 + \dots + y_{n-1} \right) \tag{51}$$

using the trapezoidal rule to determine the area under the curve for i_p . If 5-deg intervals are used in Table 1, the coefficient of Eq. (51) would be $\frac{1}{36}$.

The d-c component of grid current I_c can be found in a similar manner by substituting as ordinates the items of line 8 in Eq. (51).

The maximum amplitude of the fundamental component of the plate current is given by

$$I_{p1} = \frac{2}{\pi} \int_0^\pi i_p \cos \theta d\theta = \frac{1}{9} \left(\frac{y_0'}{2} + y_1' + y_2' + \dots + y_{n-1}' \right) \tag{52}$$

using the trapezoidal rule to evaluate the definite integral. If 5-deg intervals are used in Table 1, the coefficient of Eq. (52) becomes $\frac{1}{18}$.

The maximum amplitude of the fundamental component I_{g1} of the grid current can be obtained in the same way by substituting the items of line 10 in Eq. (52).

Table 1

Given data		Assumed values		Computed values			Equation
Tube		E_p min		E_g			48
μ		e_g max		E_c			49
E_b		θ_1		E_0			50
1	θ	0°	10°	20°	30°	40°	θ_1
2	$\cos \theta$	1	0.9848	0.9397	0.8660	0.7660	
3	$E_0 \cos \theta$						
4	$e_p = E_b - E_0 \cos \theta$						
5	$E_g \cos \theta$						
6	$e_g = E_g \cos \theta - E_c$						
7	i_p	y_0	y_1	y_2	y_3	y_4	0
8	i_g0
9	$i_p \cos \theta$	y_0'	y_1'	y_2'	y_3'	y_4'	0
10	$i_g \cos \theta$	0

19. Power Relations. The d-c power supplied to the circuit from the source of E_b is

$$P_{input} = E_b I_b \tag{53}$$

The power output to the tank circuit at the fundamental frequency is

$$P_{tank} = \frac{E_0 I_{p1}}{2} \tag{54}$$

since the tank impedance is of the nature of a pure resistance R_b at resonance. The required value of R_b is evidently

$$R_b = \frac{E_0}{I_{p1}} \tag{55}$$

and is related to the constants of the tank circuit by

$$R_b = \frac{L_0}{C_0 R_0} \tag{56}$$

where R_0 is the apparent resistance of the tank coil and includes coupled resistance introduced by the useful load which is either inductively or conductively coupled to the tank coil. In the circuit of Fig. 26 the value of coupled resistance reflected into the tank coil would be the power absorbed from the tank divided by the square of the oscillatory tank current.

The resistance of the load required to fulfill the assumed operating conditions, as given by Eq. (55), will bear no simple relation to the plate resistance r_p of the tube as used in computations relating to class A power amplifiers, since r_p is infinite during the greater portion of the cycle under class C operation. From Eq. (56) it is seen that

load impedance of the tank circuit may be varied by varying the ratio of L_0 to C_0 . As the latter item is often a mica capacitor of fixed value, a variation may be made in the value of R_0 by using the tank inductance as an autotransformer, as illustrated in Fig. 28. The ratio of transformation will be approximately the turns ratio P/S , and, by moving the plate tap so as to alter the number of turns included in P , it is possible to change the load impedance as viewed from the tube by the square of the transformation ratio.

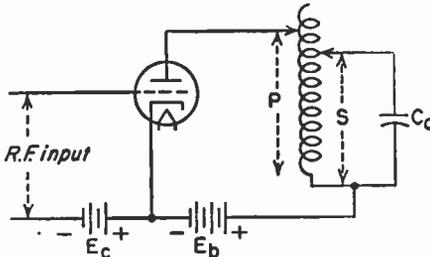
The power input to the grid is

$$P_{\text{grid input}} = \frac{E_g I_{g1}}{2} \tag{57}$$

The power amplification will be Eq. (54) divided by Eq. (57) and is

$$A_p = \frac{E_o I_{p1}}{E_g I_{g1}} \tag{58}$$

Power amplifiers are practically always operated with a fixed-bias voltage E_c instead of being self-biased by means of a grid leak and capacitor, as with oscillators. This is because in the event of failure of the excitation voltage the self-bias would no longer function and the tube would be injured. A portion of the power input to the grid would be consumed across E_c and would charge the bias battery, if one were used. This power lost across the bias is $E_c I_c$, and the power consumed within the tube due to the flow of grid current is



$$P_g = \frac{E_g I_{g1}}{2} - E_c I_c \tag{59}$$

FIG. 28. Tank-circuit inductance used as autotransformer to vary load impedance.

Since the grid is enclosed by the plate, the heating of the grid by P_g must be radiated by the plate in addition to its own losses.

The power loss within the tube which is to be dissipated at the plate in the form of heat, exclusive of the power loss in the filament, is

$$\text{Tube loss} = E_b I_b - \frac{E_o I_{p1}}{2} + \frac{E_g I_{g1}}{2} - E_c I_c \tag{60}$$

This expression may be used to check the assumed operating conditions from the standpoint of allowable plate dissipation.

The plate efficiency is defined as the ratio of the output to the tank circuit to the power supplied to the plate and is given by

$$\text{Plate efficiency} = \frac{E_o I_{p1}}{2 E_b I_b} \tag{61}$$

With the allowable plate dissipation fixed, a moderate improvement in the plate efficiency will materially increase the useful output, and the maximum output will be obtained when the plate efficiency is made a maximum.

The effective value of the oscillatory current in the tank will be

$$I_L = \frac{E_o}{\sqrt{2(R_0^2 + \omega^2 L_0^2)}} \tag{62}$$

Where the effective value of Q for the coil is high, the currents in the coil and capacitor are approximately the same and will be given with sufficient accuracy for most purposes by

$$I_L = I_c = E_o \omega C_0 = \frac{E_o}{\omega L_0} \tag{63}$$

The preceding discussion has been based upon the series-fed circuit of Fig. 26, but the same equations and method of analysis will likewise apply to the case of parallel feed in Fig. 23. This latter arrangement is the one usually employed.

20. Class B Amplifiers. In order not to distort the envelope of the applied modulated wave in Fig. 24, the dynamic characteristic must be essentially linear, and the operating conditions are chosen so as to bring this about. When this is the case, the maximum amplitude of the fundamental component of the plate current is given by

$$I_{p1} = \frac{\mu E_o}{2r_p + R_b} \tag{64}$$

to a fair degree of approximation. The d-c component of plate current will then be

$$I_b = \frac{2}{\pi} I_{p1} = 0.637 I_{p1} \tag{65}$$

The plate efficiency, from Eq. (61), becomes

$$\text{Plate efficiency} = \frac{E_o I_{p1}}{2 E_b I_b} = \frac{\pi E_o}{4 E_b} \tag{66}$$

Since E_o approaches E_b as a limiting value, it follows that the plate efficiency of a class B amplifier approaches 78.54 per cent as a limiting value. In actual practice it is

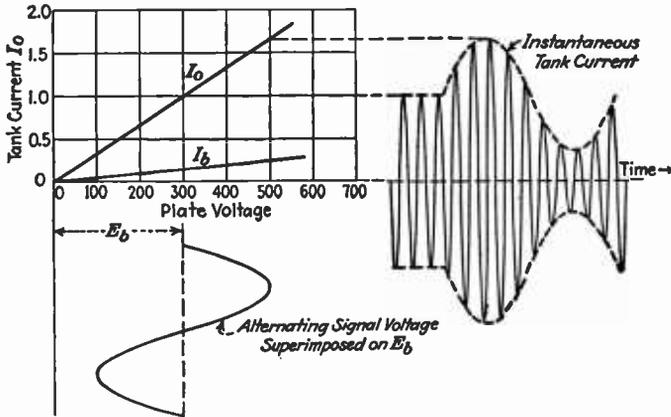


FIG. 29. Modulation of class C amplifier by superimposing a-f signal voltage on plate-supply voltage.

usually about 65 per cent on excitation peaks at 100 per cent modulation and falls to about 33 per cent when the applied excitation voltage is unmodulated.

21. Tuning Adjustments. The tank circuit should always be adjusted to unity power factor so that minimum plate voltage may coincide with maximum plate current. A departure from this relation will lower the plate efficiency. This adjustment is usually made by tuning the tank circuit for minimum d-c plate current. Strictly speaking, minimum I_b may be used as an accurate measure of unity power factor only when C_o of the tank is the element varied. The usual tuning adjustment is L_o , which is varied by means of a copper or aluminum disk rotated within the tank coil and acts as a single short-circuited turn. In this case maximum impedance will not occur at unity power factor, and L_o should be adjusted to a value slightly lower than that which produces minimum I_b . If the effective value of Q for the tank is fairly high, the adjustments for maximum impedance and unity power factor practically coincide, in which case the current may be adjusted for minimum plate current with either tuning element the variable.

22. Modulated Amplifiers. If an a-f voltage is superimposed upon the d-c plate-supply voltage E_b of a class C amplifier having constant r-f excitation applied to its grid, the tank current I_o may be made to rise and fall in amplitude as illustrated in Fig. 29. The schematic diagram of the circuit is shown in Fig. 30. A linear relation must exist between tank current and plate voltage if distortion is to be avoided. The relation between the plate voltage and I_b should also be fairly linear so that the modulator tube supplying the a-f power shall work into a constant load resistance, which will be equal to E_b/I_b , or, in general, $\Delta E_b/\Delta I_b$.

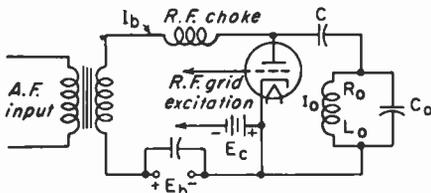


FIG. 30. Plate-modulated class C amplifier.

operation in order to avoid excessive plate heating on modulation peaks. The grid bias E_c required is approximately twice the value of cutoff for the tube, and the tank impedance is usually higher than with unmodulated operation. The plate efficiency is lower than with unmodulated amplifiers and is usually in the neighborhood of 60 per cent, depending upon the size of the tube used. Either triodes or tetrodes may be used.

The continuous power output with 100 per cent modulation is 1.5 times the power at zero modulation. The output on modulation peaks will be four times the unmodulated carrier output. This increase in the power output when modulated must be

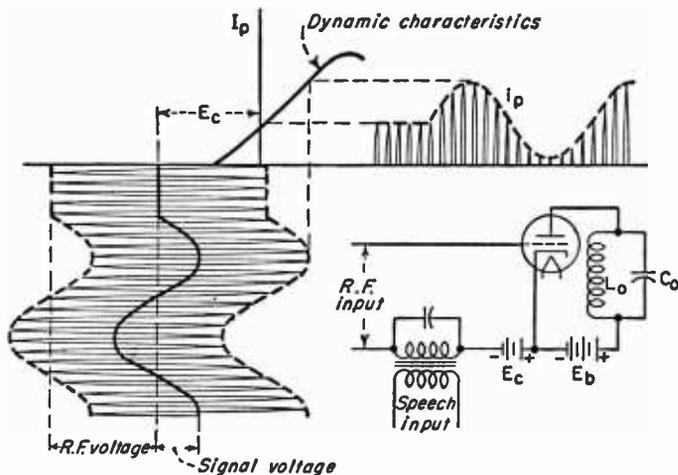


FIG. 31. Circuit and operation details of grid-modulated amplifier.

furnished by the a-f input from the modulator tubes. The amount of a-f power required varies with the square of the modulating factor, so that the modulator tubes must be capable of furnishing a sizable amount of audio power if 100 per cent modulation is to be attained.

While the plate-modulated amplifier has been widely used, other methods requiring very much less audio power can be employed. Instead of varying the voltage applied to the plate of the modulated amplifier, it is possible to secure similar results by varying

the magnitude of the C bias at an a-f rate. The schematic circuit is shown in Fig. 31, together with the details of operation. The signal voltage cyclically adds to and subtracts from the fixed biasing voltage E_c , causing the amplitude of the plate-current impulses to rise and fall. The plate-current wave shapes will be similar to those of the class B amplifier of Fig. 24, except that the angle $2\theta_1$ during which plate current flows will vary with the modulation. The mode of operation changes from an underexcited class C amplifier when unmodulated to a class B amplifier on modulation peaks, assuming complete modulation. The advantage of this method over plate modulation is that very little a-f energy is required for complete modulation. The modulating source is required to furnish only a portion of the grid-excitation losses of the amplifier in this case. The plate efficiency is somewhat lower, and freedom from distortion is more difficult to secure.

Another method is to insert the modulating voltage in the suppressor-grid circuit of a screen-grid type of power pentode operating as a class C amplifier, as shown in Fig. 32. The suppressor grid is biased negatively by a moderate amount and swings positive on modulation peaks, during which time suppressor-grid current flows. The power represented by this current has to be furnished by the modulating source, but it is negligible in comparison to the demands of a plate-modulated amplifier. The distortion is low with moderately high percentages of modulation but becomes appreciable at 100 per cent.

23. Doherty High-efficiency Amplifier.¹ The plate efficiency of a class B amplifier varies between about 33 and 65 per cent from zero to 100 per cent modulation, resulting in a rather low all-day efficiency in view of the average per cent modulation of a

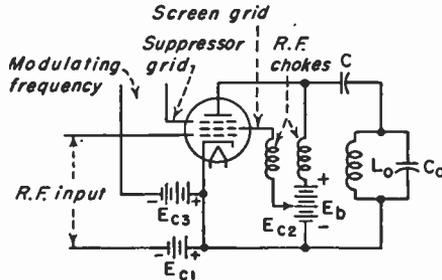


FIG. 32. Screen-grid pentode used as class C amplifier.

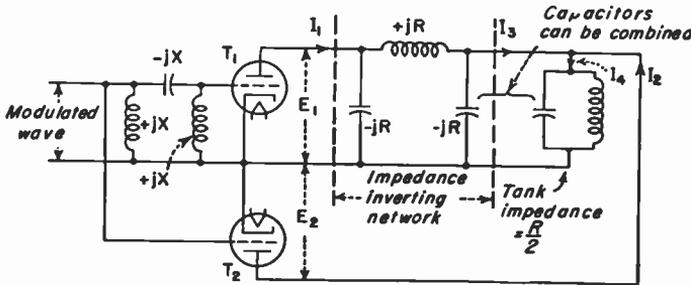


FIG. 33. Doherty amplifier.

broadcast station. Consequently an appreciable reduction could be effected in the energy requirements of a transmitter if this efficiency could be raised and kept constant. The Doherty amplifier accomplishes this desirable result in the following manner:

Two tubes, effectively in parallel, supply power to a common tank circuit, as shown schematically in Fig. 33. Tube T_1 is operated so that its output voltage E_1 is at its maximum permissible value when the unmodulated carrier voltage E_c is applied to the grid. The grid bias on T_2 is made sufficiently negative so that the output current I_2 is about zero at this value E_c of the carrier voltage. This high value of E_1 with an

¹ Doherty, W. H., A New High Efficiency Power Amplifier for Modulated Waves, *Proc. IRE*, September, 1936, p. 1183.

excitation voltage of E_c impressed is brought about by having the tube work into a load impedance of $2R$, or twice the value of tank impedance that would be ordinarily used. This value of $2R$ is brought about by the properties of the impedance-inverting network in the plate circuit, which is the equivalent of a quarter-wave line. These lines have a sending-end impedance Z_s which is given by the relation

$$Z_s = Z_L^2/Z_r \quad (67)$$

where Z_r = terminating impedance ($R/2$)

Z_L = characteristic impedance of the line, which is equal in this case to R when the reactive series and shunt arms of the simulating network have the values given in the figure

When the carrier voltage increases to a value greater than E_c (reaching $2E_c$ at 100 per cent modulation), T_2 begins to furnish power to the tank circuit. This causes the impedance of the tank, as viewed from the opposite end of the network, to rise. But this apparent rise in Z_r , from Eq. (67), causes a reduction in Z_s . Consequently, the output current I_1 of tube T_1 rises, even though E_1 remains constant. The increasing grid excitation maintains E_1 as the plate load impedance Z_s falls. As the excitation increases beyond the unmodulated amplitude E_c , T_2 contributes more and more power to the tank and thereby permits T_1 also to supply more power. When the excitation reaches a value of $2E_c$, corresponding to the instantaneous peak of a completely modulated wave, half of the power in this tank is being contributed by T_2 . The network

at that instant is effectively terminated in R ohms instead of the original value of $R/2$, permitting T_1 to deliver twice its initial power output. The total power delivered to the tank circuit is then the required value of four times the unmodulated value.

The variations in the various currents in Fig. 33 are shown in Fig. 34a, and the voltages E_1 and E_2 vary as shown in Fig. 34b. One of the characteristics of the impedance-inverting network shown in Fig. 33 is that the

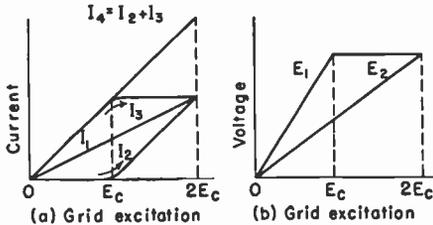


Fig. 34. Current and voltage relations in Doherty amplifier.

current I_3 will lag 90 deg behind E_1 . Consequently a network producing a similar phase shift, but in the opposite direction, is inserted in the grid circuit of T_1 , so that the currents I_2 and I_3 will be in phase with each other. Reversing the signs of the reactances in the series and shunt arms of the grid network, as shown, will produce the desired leading phase shift of 90 deg.

The Doherty method of operation enables a plate efficiency of slightly more than 60 per cent to be secured when the carrier is unmodulated. The distortion is somewhat greater than with class B operation, Fig. 34, depicting the ideal characteristics, but by using reversed feedback it is possible to meet all the requirements of high-fidelity broadcasting.

24. Frequency Multipliers. The plate current of a class C amplifier is badly distorted and contains a large percentage of harmonics. It is possible to resonate the tank circuit to one of these harmonics and cause it to absorb power at the harmonic frequency. The impedance offered to the fundamental and the balance of the harmonics will be small; hence little power will be absorbed at these frequencies.

Frequency multipliers are used to obtain higher frequencies than can be readily produced by crystal-controlled oscillators. Quartz crystals for high frequencies become rather fragile and are apt to crack in service. To secure crystal control of the frequency in the case of short-wave transmitters, the crystal is ground to oscillate at some l-f multiple of the transmitted frequency. The output of the crystal-controlled oscillator is then impressed on one or more amplifiers connected in cascade and adjusted to multiply the frequency. The usual practice is to double the frequency

with each stage, and, while greater multiplications than this can be obtained, the output falls off rapidly as higher multiplications per stage are attempted. If a push-pull circuit is being used as the frequency-multiplying stage, the output tank circuit will have to be tuned to the third harmonic of the input voltage, since even harmonics will cancel in the output circuit. A class C amplifier having a plate efficiency of 80 per cent would show an efficiency of about 70 per cent when used as a frequency doubler. The instantaneous current and voltage relations in a frequency doubler will be similar to Fig. 27 except that the frequency of e_p will be twice as great and will, therefore, be low in value for a shorter time interval. This requires a smaller value of θ_1 to keep the losses within the tube small. — These losses are proportional to the product of the instantaneous values of e_p and i_p and can be minimized by restricting the flow of plate current to a smaller interval of time. This calls for values of E_p and E_c somewhat higher than with the conventional type of class C amplifier. Either triodes or tetrodes can be used. The former will not need to be neutralized, as the input and output circuits are tuned to different frequencies and hence will not oscillate.

CHAPTER 10

VACUUM-TUBE OSCILLATORS

BY ROBERT I. SARBACHER¹

AND

DANIEL C. FIELDER²

1. Classification of Oscillators. A vacuum-tube oscillator is usually defined as a device which converts power obtained from a d-c source into alternating power. Some of the principal types of vacuum-tube oscillators and the sections in which they are discussed are listed below:

- Feedback oscillators (2)
- Magnetostriction oscillators (9)
- Resistance-capacitance oscillators (15)
- Negative-resistance oscillators (16)
- Beat-frequency oscillators (heterodyne) (22)
- Relaxation oscillators (23)
- Positive grid oscillators (Barkhausen-Kurtz) (34)
- Magnetron oscillators (36)
- Klystron oscillators (37)
- Mechanical-electronic oscillators (38)

It is customary to classify oscillators in two groups. The first group is characterized by a definite frequency and by nearly sinusoidal voltage. When such a system is started from rest, it will complete a large number of oscillations before reaching the steady state in which each cycle is identical with the preceding one. The members of this group may be called *harmonic oscillators* and include all the members of the above list except the relaxation oscillators.

Characteristics of the second group are rather indefinite frequency and extremely nonsinusoidal operation. When systems of this type are started from rest, they may reach the steady state in a very few cycles. Such oscillators are referred to as *relaxation oscillators*.

The harmonic oscillators which comprise the first group find wide application in communication circuits. They are used in timing and trigger circuits and in the generation of sinusoidal oscillations. It should be noted that the degree of tube gain and negative resistance, as well as the circuit configuration, determines to a considerable extent whether harmonic or relaxation oscillations exist. In general, the frequency of the relaxation oscillator is not very definitely fixed by the circuit elements and so is relatively easily controlled by an external influence. However, some modern relaxation oscillators have rather good frequency stability and are used without external control.

A system will not oscillate unless the various elements are properly proportioned, even if the configuration is correct. Fundamentally this means that, unless as much energy is delivered to the oscillatory circuit through the tube as is dissipated in each cycle, the oscillations cannot be maintained. For any system to oscillate stably at a definite amplitude, it is necessary that it involve some nonlinearity.³ In some cases nature of the nonlinearity is not obvious, but the effect is always there. The source of

¹ National Scientific Laboratories, Washington, D.C.

² Georgia Institute of Technology, Atlanta.

³ Le Corbeiller, P., *J. IEE, Wireless Sec.*, 11, 292, 1936.

the nonlinearity may be in the tube, in the resonant circuit, or in a special control circuit. In any system in which the tube itself is nonlinear, the stabilization is necessarily accompanied by the generation of harmonic currents and voltages, although the effect of these may be reduced by highly selective resonant circuits.

When extremely accurate frequency control is required, low-powered oscillators are used because it is then less difficult to meet the conditions required by a high degree of

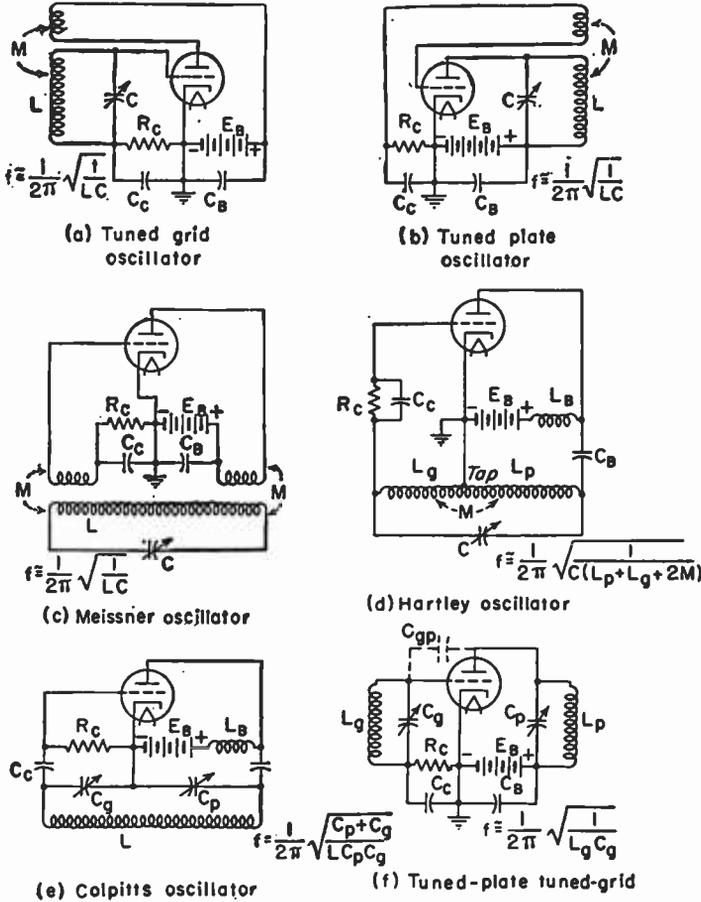


FIG. 1. Types of feedback-oscillator circuits.

frequency stabilization. One or more buffer amplifiers may be used under these circumstances to meet the power requirements of the particular application. When frequency stability is not particularly important, high-power oscillators may be used, with which tube efficiencies approaching 90 per cent may be obtained.

2. Feedback Oscillators. Oscillations may be generated with an amplifier that is connected so as to supply its own input voltage in the correct phase and magnitude. This is possible since the power required to supply the input voltage to the amplifier tube is much less than the amplified output. Oscillators operating in this way may be classed as feedback oscillators. Circuits which may be used for this purpose are shown in Fig. 1. It can be shown that, in general, the alternating voltage fed back to

the grid of the oscillator tube should be 180 deg out of phase with the alternating voltage across the plate terminals of the tube. The voltage fed back to the grid must further have an amplitude sufficient to develop the output power necessary to maintain this voltage. In the tuned-plate, tuned-grid, and Meissner oscillators, Figs. 1a, 1b, and 1c, this is achieved through mutual induction between the plate and grid circuits. In the Hartley and Colpitts oscillator circuits, Figs. 1d and 1e, the grid voltage is obtained by applying a portion of the voltage developed in the resonant circuit to the grid. In the tuned-plate-tuned-grid circuit, Fig. 1f, the energy necessary to develop the grid voltage is fed back to the grid circuit through the plate-grid capacitance of the tube.

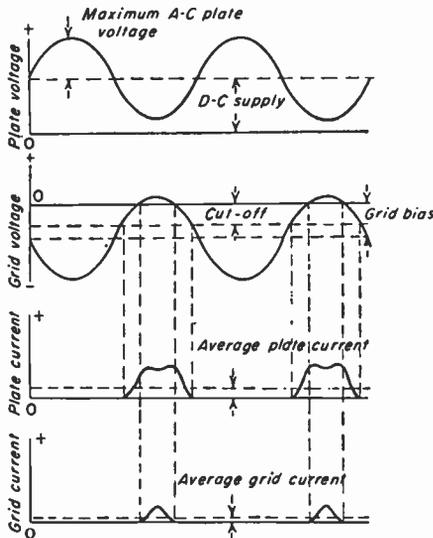


FIG. 2. Oscillator voltage and current relations.

The frequency at which oscillation occurs approximates very nearly the frequency of the resonant circuit associated with the oscillator. In the case of the Meissner and tuned-plate-tuned-grid oscillators it may be shown that, since these circuits have more than one resonant branch, they may operate at either of two frequencies when the coupling between the two circuits exceeds a critical value.

One of the most popular oscillator circuits is the Hartley. This popularity is due partly to the fact that the criterion of oscillation is not at all critical. The amplitude of oscillation is easily controlled by adjustment of the tap on the oscillator coil. For the generation of low audio frequencies, with good wave form, the Hartley oscillator is particularly suitable. This is because the resonant circuit capacitor shunts both the coils L_p and L_g and hence gives a lower frequency of oscillation for a given total inductance than either the tuned-plate or tuned-grid oscillators.

The Colpitts oscillator is less convenient to operate as a variable frequency oscillator since it is necessary to vary both C_g and C_p to maintain oscillations. However, with this type of oscillator the impedance of both the plate and grid circuits to harmonics is quite low since these circuits are shunted by C_g and C_p , respectively. This low-impedance path for the harmonic currents results in a reduction in the harmonic voltages generated in the system and hence improves the wave form. The same advantageous low impedance to harmonics also appears in the Hartley circuit if the coefficient of coupling between the two coils approaches unity.

Any of these fundamental oscillator circuits may be modified to employ two tubes in push-pull or in parallel. With parallel operation, parasitic oscillation which may be developed must be suppressed. With push-pull operation the harmonic content is decreased and the frequency stability increased over that of the single-tube circuits. Push-pull operation of oscillators is particularly advantageous at high frequencies.

When the plate supply voltage is connected in series with the plate inductances, the connection is called *series feed* (see Figs. 1a, 1b, 1c, 1f). When the plate-supply voltage is connected through a choke coil to the plate of the oscillator tube and the oscillating circuit is connected through a blocking capacitor to the plate, the connection is called *parallel feed* (see Figs. 1d, 1e). In practice, it is usually desirable to employ parallel feed since with this type of connection the resonant circuit is isolated from the d-c supply voltage.

Fixed bias is rarely used in feedback oscillators. Resistance bias, as shown, is

almost always used in order that the oscillator be self-starting and that stable operation, as is discussed in Sec. 14, be ensured.

To obtain a voltage for self-bias, it is necessary that the grid be driven positive for a portion of the cycle. The resulting grid current develops a d-c voltage across the bias resistor. Plate and grid voltage and current curves are shown in Fig. 2 for a typical feedback oscillator, such as the tuned-plate oscillator (Fig. 1b). The curves of Fig. 2 are similar to those of a class C amplifier, since the usual feedback oscillator operates as a class C amplifier. The feedback circuit fixes the ratio of a-c plate voltage to a-c grid voltage.

3. Frequency Stabilization. 1. *Causes of Frequency Variation.* There are two major causes which contribute to undesired frequency variation.¹ These are a result of changes in (1) tube characteristics and (2) circuit parameters.

Changes in the tube characteristics result in general from changes in (1) plate potential, (2) grid potential, (3) filament potential, (4) filament emission due to causes other than (3) (such as disintegration of the filament), (5) changes in spacing of tube elements, and (6) interruptions (keying) of the circuit.

Changes in the values of circuit parameters result from (1) changes in temperature of inductances, (2) changes in temperature of capacitances, (3) changes in power taken from oscillator, and (4) changes in the mechanical arrangement of the circuit elements caused by vibration, electromagnetic force, electrostatic force, and temperature.

2. *Methods of Preventing Frequency Variation.* The plate and grid polarizing potentials may be stabilized by employing voltage-regulating devices. Since the oscillator tube is usually operated so that there is an abundant space charge in the neighborhood of the filament, slight variations in heater voltage and cathode emission have a small effect. The spacing of the electrodes, which may vary slightly with tube temperature, affects the interelectrode capacitance. This effect may be minimized by the choice of a larger capacitance in the resonant circuit and the use of circuits in which the resonant circuit capacitance shunts the capacitance between the plate and grid. At higher frequencies, where the resonant circuit capacitance becomes of the same order of magnitude as the plate-grid capacitance, this effect is increased.

Changes in the values of circuit parameters, such as those caused by temperature variation of inductance and capacitance, can be reduced by temperature-controlled compensating inductances and capacitors.²

It can be shown that the frequency of oscillation will be affected by changes in load unless the power output can be taken from the system without changing the current in the inductance.³ The use of buffer amplifiers or electron coupling makes it possible to prevent changes in load from affecting the frequency. Electron coupling is discussed in Sec. 41.

By careful mechanical and electrical design it is possible to reduce the effects caused by vibration.

When a high degree of frequency stability is required in an oscillator, it is usually better to make the frequency substantially independent of the tube constants rather than try to prevent variation of these constants. The principal methods of doing this are by the use of the following:

1. Piezoelectric crystals
2. Magnetostriction rods
3. Selective filters
4. Resistance stabilization
5. Reactance stabilization
6. Bridge stabilization

4. Piezoelectric Crystal Oscillators. Oscillators which have the highest frequency stability are those which are controlled by crystals. This control is based upon the piezoelectric effect, which is a means by which a mechanical motion is coupled to an

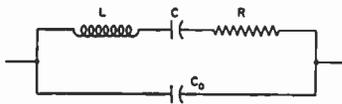
¹ Gunn, R., *Proc. IRE*, **18**, 1560, 1930; Llewellyn, F. B., *Proc. IRE*, **19**, 2063, 1931.

² Gunn, R., *Proc. IRE*, **18**, 1565, 1930; Griffiths, W. H., *Wireless Eng.*, **11**, 234, 1934.

³ Reich, H. J., "Theory and Application of Electron Tubes," p. 332, McGraw-Hill, 1944.

electric circuit. When a piezoelectric material is compressed or stretched in certain directions, electric charge appears on the surfaces of the material that are perpendicular to the axis of strain. Conversely, when such a material is placed between two metallic surfaces and a potential difference applied to them, mechanical strains are set up within the crystal. The amplitude of the voltage produced by mechanical strain may vary from a fractional voltage to several hundred volts.

There are a number of crystalline substances which exhibit this piezoelectric effect; among them are quartz, Rochelle salts, and tourmaline. Of these, quartz is used almost exclusively for controlling the frequency of oscillators because it is mechanically



Note: R is very small compared to the reactance of either circuit elements

FIG. 3. Equivalent circuit of a crystal.

the crystals vibrate is determined principally by their physical dimensions. Sections 5, 6, 7, and 8 of this chapter describe the crystals, crystal cuts, methods of temperature control, and methods of mounting.

There are many circuits in which piezoelectric crystals are employed; two representative circuits are shown schematically in Fig. 4; these have been designed by Pierce.¹ Others are described in more detail in later sections. Analysis of circuits containing crystals is greatly simplified by replacing the crystal with its equivalent electric circuit.² This equivalent circuit represents the crystal as a parallel combination of a series resonant LRC circuit and a capacitance. The equivalent circuit is shown in Fig. 3.

Care must be taken in the choice of the biasing resistor R_c , shown in the circuits of Fig. 4, since the resistance, in addition to its function of controlling the grid polarizing potential, also controls the a.c. which flows through the crystal. If this current exceeds the safe operating value for the crystal,³ the crystal may vibrate so violently as to shatter itself.

If the plate circuit of Fig. 4a is inductive, the effective input conductance of the tube is negative, and oscillations may be set up in a resonant circuit connected between grid and filament. To keep the plate resonant circuit inductive, it must be tuned to a frequency slightly higher than that of the crystal.

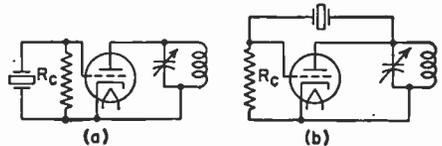


FIG. 4. Types of piezoelectric-controlled oscillators.

In the circuit of Fig. 4b, the crystal is connected between the plate and grid of the oscillator tube. This circuit will oscillate only when the plate circuit is capacitive, and hence the natural frequency of the plate resonant circuit must be slightly lower than that of the crystal.

The resonant curve of a crystal is extremely sharp, and it is this characteristic of the crystal that makes it suitable for use in controlling the frequency of oscillators. The standard measure of the sharpness of resonance of a crystal of an electrical circuit is usually denoted by Q and is numerically equal to the ratio of the total inductive reactance to the total effective resistance of an oscillating circuit. The selectivity Q of the equivalent circuit of a crystal is of the order of magnitude of 100 times that

¹ Credit is due G. W. Pierce on many crystal oscillator circuits which have been accredited to others. See his patents U.S. 1789496, filed February, 1924, and U.S. 2133642 through U.S. 2133648 filed between 1926 and 1931.

² Van Dyke, K. S., *Proc. IRE*, **16**, 742, 1928; and Mason, W. P., *Proc. IRE*, **23**, 1252, 1935.

³ In general, the safe operating value for the current through the crystal may be set approximately at 100 ma for l-f crystals and about one-half this value for crystals operating above 1 Mc.

which can be attained with ordinary inductances and capacitances. In view of this high selectivity, the crystal can oscillate over only a very narrow frequency range. With temperature-controlled crystals, frequency variations of as little as ± 2 parts in 10^7 are not uncommon. With a special bridge circuit developed by L. A. Meacham, described in Sec. 13, short-time frequency drift may be kept within ± 6 parts in 10^{10} .

The output of crystal oscillators may vary from a fraction of a watt to several hundred watts. In applications where extremely constant frequency is required, the oscillators are usually designed for low power output, and one or more buffer amplifiers are used. In this way, the crystal current may be kept small and the heating effects due to it minimized. The buffer amplifier also greatly reduces the effect on the oscillator of variations in load. With modern high-gain pentodes, operating in crystal-controlled circuits, reasonably good frequency stability at high power output may be obtained. This stability is usually sufficient for the requirements of amateur communication.

The frequency of negative resistance oscillators may also be controlled by the use of crystals.¹

5. Piezoelectric Crystals.² The occurrence of quartz crystals (the most commonly used of the piezoelectric materials) in the natural state is quite generally known. These crystals, while rarely symmetrical in form, have the general shape of a hexagonal prism, sometimes surmounted on the ends by a hexagonal pyramid. A cross section of a symmetrical crystal is shown in Fig. 5. In this diagram, the electric axes (so called because the greatest piezoelectric activity is observed in the direction of these axes) are represented by the lines XX , $X'X'$, and $X''X''$. The other axes, YY , $Y'Y'$, and $Y''Y''$, have been given the name "mechanical axes."

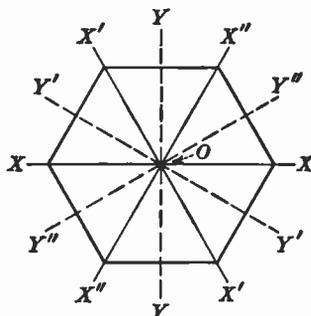


FIG. 5. Quartz crystal cross section.

Through the point O , perpendicular to the plane of the page, passes the optic axis (Z axis) of the crystal. Sections or plates are cut from the crystal for use as highly selective circuit elements.

Crystals cut perpendicular to the X axis are called X cut, and crystals cut perpendicular to the Y axis are called Y cut or 30-deg cut (see Fig. 6). Although both the X and Y cut have been used extensively, they are now largely superseded by more modern cuts which greatly improve the performance of crystals.

6. Piezoelectric Crystal Cuts. One of the objections to the X - and Y -cut crystals is their large temperature coefficient, amounting to -10 to -25 parts per million per degree centigrade for the X cut, and $+100$ to -20 parts per million per degree centigrade for the Y cut. When they are used as frequency-control elements, provision must be made to keep their temperature constant. Also, these plates often exhibit discontinuous frequency-temperature characteristics. This characteristic of the Y -cut plate can be improved by suitable grinding, while the X -cut plate cannot be improved, and may often be inoperative at the desired frequency of operation.

From the statement above, regarding the range of the temperature coefficient for the Y -cut plate, it might seem possible to get a plate having a zero temperature coefficient. Morrison³ found this to be the case for the so-called *ring* or *doughnut plate* when operated at a temperature of approximately 40°C (see Fig. 8). This plate is, however, very difficult to grind and, therefore, expensive. Moreover, it exhibits a number of spurious resonances near the desired frequency.

More recent work has resulted in the discovery of a number of plates which overcome most of the difficulties encountered with those plates mentioned above. These plates

¹ MacKinnon, K. A., *Proc. IRE*, **20**, 1689, 1932.

² See references at end of section.

³ Morrison, W. A., *Proc. IRE*, **17**, 1103, 1929, and *Bell System Tech. J.*, **8**, 493, 1929.

are obtained by cutting the crystal in such a way that at least two faces of the plate are not perpendicular to the crystallographic axes. Some of these plates are considered below.

Two cuts which are suitable for operation above 500 kc are the *AT* and *BT* cuts. These have a zero temperature coefficient when operated at temperatures of approximately 45 and 25°C (see Fig. 8).

The *AT* plate is obtained by cutting the crystal at an angle of rotation about the *X* axis of 35 deg (see Fig. 7). The *BT* plate is obtained at an angle of -49 deg as indicated.

When the thickness of the *AT* and *BT* plates is increased to obtain lower operating frequencies (below about 500 kc), difficulties arise due to coupled modes of vibration.

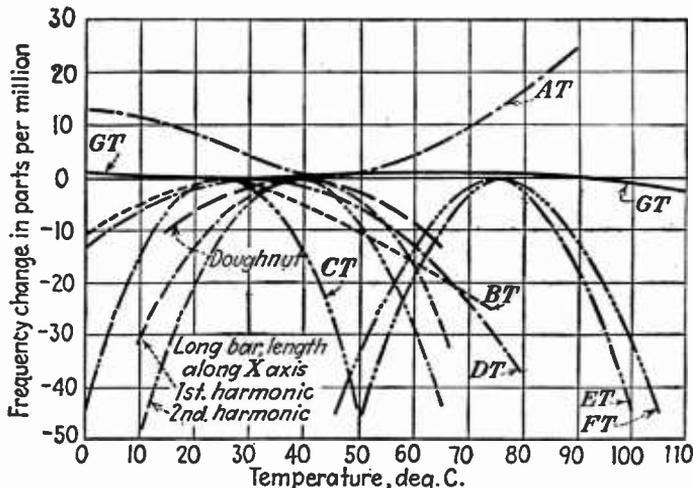


FIG. 8. The temperature coefficient of frequency for different crystal cuts. (Courtesy of Bell System Tech. J.)

Even though elastic coupling between desired and undesired modes of vibration in these plates is small, it becomes important when the frequencies approach one another, as is the case when the thickness dimension becomes comparable with the other dimensions. To avoid the use of unusually large plates of quartz for lower frequency operation, two new types of quartz crystal elements have been developed. These are known as the *CT* and *DT* plates and are directly related to the h-f low-temperature-coefficient *AT* and *BT* plates. The temperature coefficient of frequency of these new plates may be made zero by operating them at a suitable temperature, as indicated in Fig. 8. These *CT* and *DT* plates are useful as stabilizing elements for oscillators operating between 50 and 500 kc.

The *ET* and *FT* crystal cuts have zero temperature coefficients at higher temperatures than those discussed above. Their useful range, which is from 100 to 1,000 kc, extends to higher frequencies than that of the *CT* and *DT* plates. This is because they operate at a harmonic of the fundamental vibration. The *GT* crystal cut,¹ has a constant frequency over a very wide temperature range. As can be seen in Fig. 8, the shape of the temperature-frequency curve is different from that of the other special cuts. The superiority of this cut, particularly when temperature control is not used, is evident. The *GT* cut is very satisfactory at frequencies between 100 and 500 kc. Other types of crystal cuts² are the -18° cut, and the 5° cut, the *NT* cut, and the *MT*

¹ Mason, W. P., *Bell System Tech. J.*, 19, 74, 1940.

² For further information in regard to crystals, see "Quartz Crystals for Electrical Circuits," R. A. Heising, Van Nostrand, 1947.

cut. The -18° cut crystal is used in filters, while the 5° cut crystal is used in filters and in oscillators, operating at frequencies as low as 5 kc. The *NT*-cut crystal may be used in a frequency range from 4 to 50 kc. The *MT*-cut crystal is used in the frequency range between 50 and 500 kc.

There is no definitely established frequency limit for quartz plates; the practical limits are being constantly extended. Quartz plates are rarely called upon to control more than a few watts directly; higher powers are controlled by amplifying the output of the crystal stage.

Several other materials which assume a more or less well-defined crystalline form have been investigated as possibilities for piezoelectric elements. Among these may be mentioned tourmaline and Rochelle salt. The Rochelle salt crystals have, in general, been discarded, although they have found applications in loud-speakers, microphones, and phonograph pickups.

Tourmaline, while it is practically as good as quartz over a great frequency range—and somewhat better than quartz in the range from about 3 to 30 Mc—has the disadvantage of being a semiprecious stone. Its cost is, in consequence, out of proportion to its usefulness.

Beyond the range where crystals exert satisfactory control, special resonant circuits of extremely high selectivity may be used as frequency-control elements (see Sec. 33).

7. Temperature Control of Piezoelectric Crystals. Since the resonant frequency of all crystals, particularly of the *X* and *Y* cuts, changes with temperature, it is necessary, if a high degree of frequency stability is required, to make some provision to keep the temperature of the crystal constant. In some cases, where every possible precaution is taken to prevent frequency variation, the associated electrical circuit as well as the crystal is maintained at a constant temperature.

Electric ovens suitable for temperature control of crystals are usually designed after the principles given by Marrison.¹ These principles involve the thermal conductivity of the material of which the oven is made, the ambient-temperature range, and the temperature coefficient of the quartz plate.

Briefly stated, the problem is one of accurately determining the temperature at which it is desired to maintain the plate and of causing any slight deviation from this temperature to actuate suitable thermostatic devices, which in turn cause more or less current to flow through the heater associated with the oven.

An example of such a control chamber is given by Marrison as follows:

It consists of a cylindrical aluminum shell with a wall about one inch thick, with a heater, and with a temperature-responsive element in the wall to control the rate of heating. The aluminum shell has a metal plug that screws into the open end forming a chamber for the crystal which is then completely closed except for a small hole for electrical connections.

Since aluminum is a good thermal conductor the shell equalizes the temperature throughout the chamber and thus avoids the use of a fluid bath. The main heating coil is wound in a single layer over the whole curved surface of the aluminum cylinder, being separated from it only by the necessary electrical insulation. Auxiliary heating coils are wound also on the ends so as to distribute the heating as uniformly as possible. This, in effect, makes the short cylinder behave like a section from an infinite cylinder. To protect the thermostat from the effect of ambient temperature gradients the heating coil has an outside covering consisting of four layers each of thin felt and sheet copper spirally wound so that alternate layers are of copper and felt, the innermost layer being of felt and the outer one of copper. . . . This covering is very effective in reducing surface gradients since the conductivity in directions parallel to, and perpendicular to, the surface differ by a large ratio.

The thermostat used with these constant-temperature chambers is generally the mercury-column type. This is essentially a thermometer in which contact wires have been fused. At the point on the scale where the operating temperature is located, the glass stem has been drawn out; *i.e.*, if the device is to function at, say 35°C , the stem of the thermometer is constricted and elongated between about 34.5° and 35.5° . One of the contact wires is fused through the glass at the 35° point, the other wire making contact with the mercury at the bulb. This elongation of the stem over a range of 1° or so

¹ Marrison, W. A., *Proc. IRE*, 18, 976, 1928; Clapp, J. K., *Proc. IRE*, 8, 2003, 1930.

causes the mercury column to move an appreciably greater distance per fraction of a degree change in temperature.

This type of regulator is very sensitive to minute temperature changes but is expensive, fragile, and cannot carry any appreciable current. For this latter reason, it is customary to utilize the regulator simply to change the grid bias on a vacuum tube; the tube plate circuit includes the winding of a relay which operates with small changes of plate current. This relay, which is generally too small to handle the heater current, actuates still another relay to open or close the heater circuit.

With the advent of the new crystal cuts, the temperature coefficient of frequency is so low that temperature control is normally not required. Some types of service, notably aircraft radio, where ambient temperatures may range from -40 to $+40^{\circ}\text{C}$, still require some kind of temperature regulation, but the requirements are satisfactorily met with a more or less conventional heating chamber and an ordinary bimetallic thermostat.

8. Mountings for Piezoelectric Crystals. In general, there are three types of mountings or holders for piezoelectric crystals: clamp type, wire type, and air-gap type.

Prior to mounting, the crystal faces are plated with a thin film of metal. A description of a typical plating process is given in a subsequent paragraph.

In the clamp type, the crystal electrodes form clamps which hold the crystal in place. The point of contact of the clamps is located at a nodal point, or, by definition, a point of zero motion. This method of mounting is satisfactory for certain l-f crystals. With the wire type of mounting, a section of wire is attached to each of the crystal faces. The other ends of the wires are attached to the terminals of the complete crystal assembly. Since the wires are rigidly attached to the crystal, they vibrate in flexure with the crystal. A solder ball is attached at a nodal point along each of the wires and its mechanical inertia restricts the vibrations to the section of wire between the solder ball and the crystal. With the air-gap type of mounting, an unplated crystal is held between two flat electrodes. An air gap of definite thickness between the electrodes and the crystal allows the crystal to vibrate between the electrodes. The outer edges of the crystal are firmly secured to prevent lateral motion. The shunting capacitance of this mounting (shown in Fig. 3 as C_0) is always larger than the capacitance of the holder alone.

While the actual construction of crystal holders is beyond the scope of this discussion, it may be pertinent to point out some of the requirements which must be met by the air-gap type of holder.

These may be enumerated briefly as follows:

1. The electrode surfaces must be lapped perfectly flat and must be entirely free of oil and dirt.
2. The electrodes must be made from metal which will not corrode.
3. Some type of construction is generally necessary which will prevent lateral motion of the plate; this may be accomplished by enclosing the plate and electrodes in close-fitting cases of suitable insulating material.
4. The entire assembly should be made dustproof and evacuated if possible.

The methods by means of which the electrodes are plated directly on the quartz are known as the *sputtering* and *evaporation* processes. Mr. H. W. Weinhart of the Bell Telephone Laboratories has prepared the following description of the technique used in these processes. He states:

Films of metal can be deposited on quartz by sputtering or evaporating on the material. Some metals sputter much more readily than others, for example, gold, silver and platinum films can be deposited at a greater rate than aluminum. Metals that sputter slowly are, therefore, usually plated on by the evaporation process.

Sputtering is a process involving the releasing of atomic particles of metal by electron and ion bombardment in a gas. The usual method, when plating with air as the gas, is to place the material on which the metal film is to be deposited in a bell jar with a vacuum pump attached. A cathode of the metal to be plated is mounted about $1\frac{1}{2}$ in. above the recipient, and a small leak valve that can be regulated is attached to the apparatus, together with an aluminum anode located in the tube connection for vacuum pumping.

The system is pumped out, with the leak adjusted to maintain a pressure of 0.06 mm of mercury. If a potential of about 1,900 volts is applied between the anode and cathode, through a suitable resistance, the gas in the chamber is ionized and the cathode is bombarded. The atomic particles of metal released from the cathode surface diffuses as a gas and a metallic film is deposited on the quartz.

Evaporation of metal for the deposition of metal films on quartz is a process in which a vacuum chamber is used that can be pumped out to maintain a pressure of 10^{-4} to 10^{-6} mm of mercury. The evaporation unit can be in the form of a wavy wire, and made from tungsten 1/0.020 in. in diameter wound in a close spiral, one-eighth inch in diameter, and then stretched to form wide pitch spiral turns.

Wire, 1/0.010 in. in diameter, of the metal to be plated, is cut into short lengths and formed into hairpin shape. One piece is placed in each depression in the tungsten wire.

When the proper pressure is attained in the vacuum system, the tungsten wire is slowly heated, until the metal to be plated is melted and flows over the wire or forms globules in the depressions. Slow heating is necessary to maintain a low pressure by pumping out the liberated gases during this part of the process. By increasing the temperature of the tungsten wire the attached metal is evaporated, and deposits by condensation on the quartz surfaces, thus forming a metal plating.

For some metals such as chromium and beryllium, the preparation of the evaporator unit differs. For plating chromium the usual procedure is to plate the wavy tungsten wire with chromium electrolytically, and then to evaporate it off. Beryllium can be attached to the tungsten wire by spot welding on small pieces along the length of the wire.

9. Magnetostriction Oscillators. Oscillators having their frequency controlled by magnetostriction rods were first described by G. W. Pierce.¹ Magnetostriction in

metals is somewhat analogous to the piezoelectric effect in crystals. There is an expansion or contraction of magnetic materials as a result of magnetization and, conversely, a change of magnetic permeability as a result of mechanical stress.

If a rod of magnetostrictive material is placed in an alternating magnetic field, the rod will vibrate longitudinally at a frequency which is twice that of the a.c. producing the field. If, however, the rod is magnetically polarized, the frequency of vibration will be that of the applied a.c. Under this condition the rod may be clamped or pivoted at its

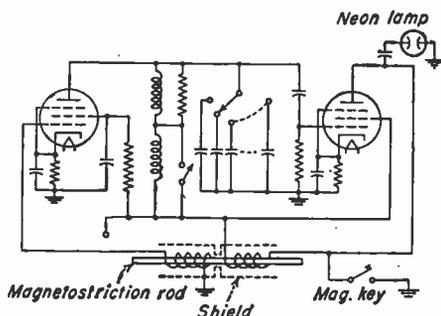


FIG. 9. Magnetostriction oscillator.

exact center, this being a nodal point. For this condition, the resonant frequency of the rod (usually in the range from one to several hundred kilocycles) is given by

$$f = \frac{v}{2l}$$

where v = velocity of sound in rod

l = length of rod

The circuit of Fig. 9 shows an improved magnetostriction oscillator.² It consists essentially of a two-tube impedance-coupled amplifier having input and output coils shielded from each other except for electromechanical coupling through the vibration of a magnetostrictive rod placed axially in both of them. A neon-glow lamp serves as an indicator of oscillation when connected across the plate coil. Operation of this circuit is dependent upon the correct choice of coupling impedance with regard to the direction of connection of the rod coils and upon the existence of good electromagnetically shielding between the two rod coils. The proper value of the coupling impedance is not at all critical since it requires practically no adjustment over a wide range of

¹ Pierce, G. W., *Proc. Am. Acad. Arts Sci.*, 63, April, 1928; reprinted in *Proc. IRE*, 17, 42, 1929.

² Pierce, G. W., and A. Noyes, Jr., *J. Acoust. Soc. Am.*, 9, 185, 1938.

frequencies. Magnetostrictive rods for use with this type oscillator have been cut accurately to length to give fundamental frequencies ranging from 5 to 60 kc.

Pierce has given extensive data on oscillators of this type, including such matters as temperature coefficients and values of the function v in the above equation for various magnetostrictive materials.

In making magnetostriction rods, nickel, Monel metal, Invar, Nichrome, Stoc metal, and other nickel alloys may be used. Because it is difficult to design magnetostriction rods which have a high natural frequency of oscillation, their use is restricted as cited above. Rods may be designed for very low frequencies by loading them at the ends or by using a tube made of magnetostrictive material which is filled with lead or other material which has a low velocity of propagation of compressional waves. Short-time frequency stabilities of 3 parts in 10^6 have been obtained with oscillators of this type without temperature control. If the temperature of the rod is kept constant, this stability may be increased. By making the rods of special alloys having a low temperature coefficient or making them of a shell of two magnetostrictive materials of opposite temperature coefficient, the change in frequency with temperature may be reduced.

10. Tuned-filter Oscillators. The tuned-filter oscillator is essentially a multistage-feedback oscillator. By feeding back the output of a highly selective multistage amplifier to the input, very good frequency stability may be obtained. Such an oscillator was described by Gunn¹ and is shown in Fig. 10. Except that the amplification takes place in more than one tube, the principle of operation of this oscillator is the same as that described under Feedback Oscillators. The frequency stability is improved by the use of more stages and by the use of more complex filter sections which have a more selective filter action. When oscillators of this type are used at radio frequencies, it is necessary to take particular care that feedback in the individual stages does not occur. Use of tetrodes and pentodes and careful shielding are necessary. Gunn gives the following data as evidence of the excellent stabilizing action. At an a.f. of 1,000 cycles, a 50 per cent change in plate potential of a two-stage system resulted in a frequency shift of less than 1 cycle. At radio frequencies a change in

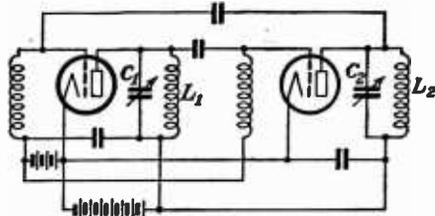


FIG. 10. Oscillator stabilized by selective filters.

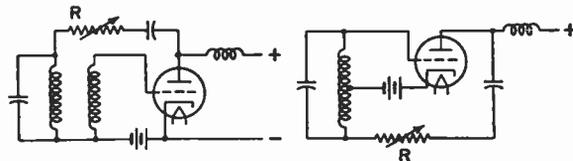


FIG. 11. Resistance-stabilized oscillator.

were taken with battery-operated filaments. If alternating filament voltage is used, the filament must be of the noninductive type. The use of a buffer amplifier between oscillator and load will improve the frequency stability.

11. Resistance Stabilization.² One of the easiest methods for improving the frequency stability of standard oscillators is by resistance stabilization (see Fig. 11). It was pointed out previously that one of the factors contributing to frequency drift is change in the plate resistance of the tube. The method of resistance stabilization consists of inserting a high resistance between the plate and resonant circuit of an oscil-

¹ Gunn, Ross, *Proc. IRE*, 18, 1560, 1930.

² Horton, J. W., *Bell System Tech. J.*, 3, 508, 1924.

plate potential of 10 per cent results in a frequency shift of 0.0003 per cent of the fundamental frequency. Changing the filament potential 8 per cent changes the frequency less than 0.0003 per cent. The above data

lator so as to make the total effective resistance of the plate circuit so high that variations in the plate resistance of the tube are relatively unimportant. This resistance also performs a second useful function. It makes a convenient means of controlling the amplitude of oscillation by controlling the feedback voltage. Obviously the power consumed by the resistance reduces the efficiency of the system.

Terman has given useful design information for this type of stabilized oscillator.¹ He recommends the following:

1. Amplification factor of tubes should lie between 4.5 and 8.
2. Turns ratio of grid and plate coils should be unity, and coupling should be as close as possible.
3. Feedback resistance should be of the order of from two to five times the plate resistance.
4. Grid bias battery must be used and not grid leak resistance.
5. For a-f oscillators, feedback resistance should not be greater than 500,000 ohms.

12. Reactance Stabilization. A more general type of stabilization than those previously presented has been worked out by Llewellyn.² He has shown that the frequency of oscillation may be made invariant to tube characteristics and hence to polar-

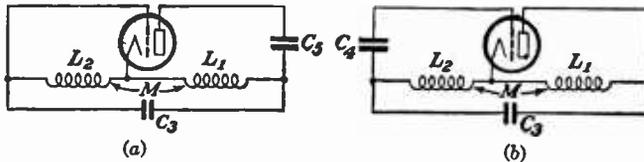


Fig. 12. Hartley oscillator. (a) Plate stabilization. (b) Grid stabilization.

izing potential by the insertion of capacitance or inductance in series with the grid or plate of the oscillator tube, or both. In his analysis, Llewellyn makes the following assumptions:

1. The resonant circuits of the oscillator have negligible losses.
2. The oscillator tube operates in a linear region of its characteristic.

He then sets up the equivalent circuits for each type of feedback oscillator and the circuit equations applied thereto. From the general solution of these circuit equations he obtains the conditions which make the frequency of oscillation invariant to the tube parameters. Representative results obtained in this way are shown in Fig. 12. For the assumption of negligible losses in the resonant circuits to hold reasonably well, it is necessary that a buffer amplifier be interposed between the oscillator and the load. This buffer stage must be very loosely coupled to the oscillator so as to draw but a very small fraction of the available power. To meet the second assumption, some form of amplitude control such as described in Sec. 14 must be used. Llewellyn further states that with unity coupling between the plate and grid circuits, the frequency of an oscillator depends only upon the inductances and capacitances in the circuit and is independent of plate resistance, grid resistance, and amplification factor, provided (1) the losses in the external circuit are small and (2) the harmonic voltages across the tube are small enough to allow the plate and grid impedance to be purely resistive.

The examples of circuit proportions in Fig. 12 will provide impedance stabilization of a Hartley oscillator, provided the assumptions made in text are met. For many more examples see F. E. Terman, "Measurements in Radio Engineering," p. 295 (1936).

When the stabilized circuits that are shown in Fig. 12 are constructed, the stabilizing inductances and capacitances may serve other functions in the circuit. For example, plate stabilizing capacitors may serve also as blocking capacitor for the plate polarizing potential. Also the grid stabilizing capacitors may serve to furnish grid bias when shunted by a high resistance. When the stabilizing capacitor is thus shunted, its

¹ Terman, F. E., *Electronics*, July, 1933, p. 190.

² Llewellyn, F. B., *Proc. IRE*, **19**, 2063, 1931; Stevenson, G. H., *Bell System Tech. J.*, **17**, 458, 1938.

required value is altered and its effectiveness reduced. The higher the resistance, consistent with the limitation discussed in Sec. 14, the smaller its effect on the required value of stabilizing capacitance. The correct stabilizing capacitance is best determined experimentally. The interelectrode capacitances of the tube are of small importance in those circuits where these capacitances form a portion of the resonant circuit. In cases where the interelectrode capacitance cannot be combined in this way, variation from predicted performance may be partly explained.

Another factor which may produce variation from the theory is the existence of harmonics as previously mentioned. The provision of a low reactance path for the harmonics will reduce their effect (see discussion on Colpitts circuit, Sec. 2).

If a variable-frequency oscillator is stabilized in this way, it is necessary to adjust the stabilizing capacitance when the frequency is varied. For this type of stabilized oscillator, at 1-Mc operation the frequency varied less than 10 cycles when the plate potential was reduced 50 per cent, and there was practically no change when the filament current was reduced 50 per cent.

13. Bridge Stabilization. The bridge-stabilized oscillator was developed by L. A. Meacham¹ and is a constant-frequency oscillator of extremely high selectivity. Short-time frequency variations no greater than ± 6 parts in 10^{10} have been obtained with a single-tube circuit.

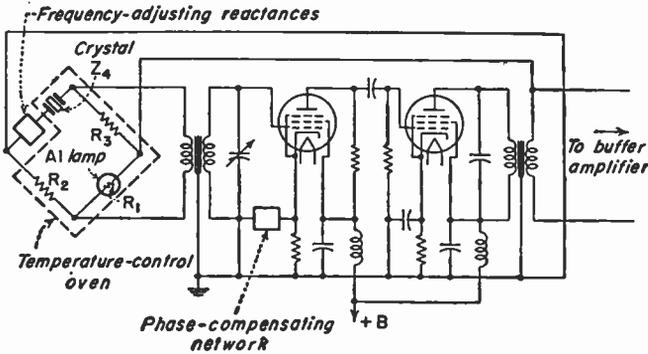


FIG. 13. Bridge-stabilized oscillator.

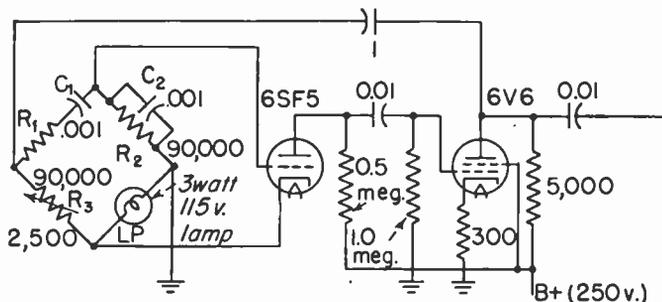
This type of oscillator, which consists of an amplifier and a Wheatstone bridge, is shown in Fig. 13.

A crystal Z_4 of high selectivity forms one of the arms of the Wheatstone bridge. Two other arms are made up of the fixed resistances R_2 and R_3 . The fourth arm R_1 is a thermally controlled resistance. The output of the amplifier is impressed across one of the diagonals of the bridge, and any unbalanced potential appearing across the conjugate diagonal is supplied to the input terminals of the amplifier. The thermally controlled resistance R_1 is a lamp and is so designed as to keep the bridge out of balance sufficiently to sustain oscillation. Since the temperature of the lamp filament is dependent upon the amplitude of oscillation, any slight variation in this amplitude or in the gain of the amplifier is immediately corrected by a small readjustment of the bridge balance. The frequency of oscillation is stabilized at that value for which the crystal impedance is purely resistive, because only at this frequency can the Wheatstone bridge approach balance. It can be shown by means of a vector diagram that a large phase shift introduced in the amplifier results in a very small frequency shift and phase shift in the crystal, owing to the phase magnifying property of a nearly balanced bridge.

¹ Meacham, L. A., *Proc. IRE*, **36**, 1278-1294, October, 1938; *Bell System Tech. J.*, **17**, 574-591, October, 1938; *Bell Lab. Record*, **18**, January, 1940.

When the polarizing potentials are supplied to the amplifier, oscillations build up rapidly since the lamp R_1 is cold and its resistance correspondingly low, resulting in low attenuation of the bridge. When the lamp filament heats up, its resistance increases and approaches the value for which the loss in the bridge equals the gain of the amplifier. If the lamp resistance exceeds its balance value, the unbalance potential becomes too small or even inverted in phase, causing the amplitude to decrease to the equilibrium value. Hence the amplitude of oscillation is also stabilized since the power required to give the lamp a resistance closely approaching that of its balance value is always very nearly the same. Variation in the amplifier gain would cause a readjustment of the tube balance, but resulting variations in the amplifier output or in the value of R_1 would be extremely small.

In place of the crystal in the Z_4 arm of the bridge a coil and capacitor connected in series could be substituted. Also a parallel resonance circuit could be used by exchang-



$$R_1 C_1 = R_2 C_2, f = \frac{1}{2\pi R_1 C_1} = 1768 \text{ c.p.s.}$$

FIG. 14. Wien-bridge oscillator.

ing its position in the bridge with R_2 or R_3 . In Meacham's bridge, Z_4 represents a crystal suitable for operation at its low-impedance or series resonance. This mode of operation minimizes the effects of stray capacitance. He has also found that a small tungsten-filament lamp of low wattage rating is quite suitable. The operating temperature of the lamp is made sufficiently high so that variations in ambient temperature do not appreciably affect balance adjustments. This temperature is found to be low enough to ensure extremely long filament life.

The use of a two-stage amplifier, as shown in Fig. 13, provides high gain and correspondingly high stability. This circuit was designed by Meacham for the Bell System Frequency Standard. Small manual adjustment of frequency is provided by the variable reactances in series with the crystal. Because of the possibility of any tendency of the circuit to break into undesired oscillation as a result of its high gain, the phase-compensating network indicated in the cathode circuit of the first tube is used.

The Wien-bridge oscillator also employs bridge stabilization to produce stable oscillations. The frequency selective Wien bridge, together with a suitable filament-type lamp, forms a stabilized frequency and amplitude control for generation of frequencies in the audio range. A Wien-bridge oscillator is used as the basic frequency generator in many commercial audio oscillators. One or more stages of amplification are required with the feedback being taken from the output of the last amplifier tube. Positive feedback is developed across the frequency selective resistance-capacitance network of the bridge, so tuned that the maximum feedback is developed at the desired frequency. In addition, the phase shift of the positive feedback voltage is zero only at the desired frequency, permitting one single frequency to be generated. The negative feedback voltage is developed across a filament-type lamp, which, by

virtue of its temperature-resistance characteristic, serves to regulate the amplitude of oscillations. The negative feedback voltage is substantially constant over a wide frequency range and is adjusted to be slightly less than the maximum positive feedback voltage. A highly stable sine-wave oscillator, with output amplitude very constant over a wide frequency range, is obtained. Figure 14 shows the schematic diagram of such an oscillator.

14. Amplitude Control. Control of the amplitude of oscillation is necessary to ensure stable operation. Also, amplitude control aids in the reduction of harmonic distortion and in the stabilization of frequency. In the feedback oscillators of Fig. 1, the amplitude of oscillation is usually controlled by the use of the grid-bias resistor and capacitor as shown. This aids in making the oscillator self-starting, for initially the bias is zero and the plate current and amplification are large. When any voltage of the frequency of the resonant circuit is set up in the system, caused by thermal agitation or transient conditions, the building up of oscillations will start. This building-up process is accompanied by the flow of grid current, which develops a direct voltage across the grid-resistor-capacitor combination, biasing the grid negatively. As the amplitude of oscillation continues to increase, the grid current increases, increasing the grid bias and decreasing the amplification of the tube. This process continues until the amplification is reduced to the point where equilibrium is established. Conversely, any decrease in the amplitude of oscillation causes an increase in the amplification and a reduction in grid bias, tending to produce stable oscillations.

If a fixed bias is used with class C operation of the oscillator, the system will not be self-starting when the plate voltage is applied since the grid bias is greater than the cut-off value.

When the time constant of the grid-resistor capacitor circuit is too large, the bias voltage adjusts itself too slowly with rapid changes in the amplitude of oscillation. This may result in a dying out of oscillations before the bias can change appreciably. When the oscillations have ceased or are about to cease, the capacitor charge leaks off through the grid resistance, and oscillations build up again to the equilibrium value. This process may repeat itself, resulting in what are called intermittent oscillations.

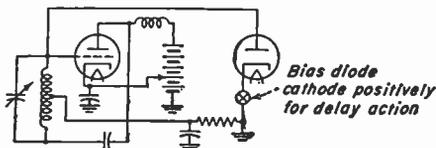


FIG. 15. Use of diode to control amplitude of oscillations.

Another method which may be used to control the amplitude of oscillation employs a diode rectifier as the limiting device.¹ This type of control is particularly suitable for oscillators operating in class A, in which no grid current flows. Figure 15 shows a Hartley circuit equipped with automatic amplitude control. The action is essentially that of a simple volume-control system employing a diode. By employing a triode, tetrode, or variable- μ pentode, the control system can also be arranged so that it does not start to operate until the amplitude has reached some predetermined level, and, in addition, the amplification introduced will increase its sensitivity. Equilibrium conditions may be obtained with small amplitudes of oscillation, where the operating conditions are substantially those corresponding to class A operation of the oscillator tube. Under these conditions very good frequency stability may be obtained, with good wave form and practically constant amplitude of oscillation as the frequency of oscillation is changed.

Other methods for controlling the amplitude of oscillation are described in Sec. 13 on the Meacham-bridge-stabilized oscillator and the Wien-bridge oscillator.

15. Resistance-Capacitance Oscillators. In this type of oscillator, shown in Fig. 16, RC networks, rather than tank circuits containing inductances, control the frequency of oscillation. The phase-shift oscillator is typical of this group. Part of the output of a single tube is fed back through three L-section filters to obtain the 180-degree phase shift required at the grid for oscillation. A minimum of three L sections is

¹ Arguimbau, L. B., *Proc. IRE*, 21, 14, 1933; Groszkowski, J., *Proc. IRE*, 22, 145, 1934.

required, since the maximum phase shift which can be obtained with a single section is less than 90 deg. The combination of the three L sections allows only one frequency to be shifted in phase by 180 deg and thus stable oscillations of this frequency exist.

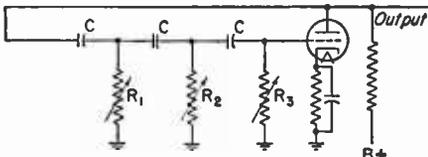


Fig. 16. Phase-shift oscillator.

Frequency may be varied by changing either R_1 , R_2 , or R_3 in Fig. 16. The wave form of this oscillator is excellent if the bias on the tube is adjusted for class A operation and the plate resistor is chosen barely to allow oscillation. The oscillator will operate with the capacitors and resistors of the phase-shift network interchanged, but the wave form obtained is not so good as

that obtained with the original configuration.

An RC audio oscillator suggested by Scott¹ is shown in Fig. 17. The inverse feedback principle is used to obtain sharply selective circuits in which inductances are not

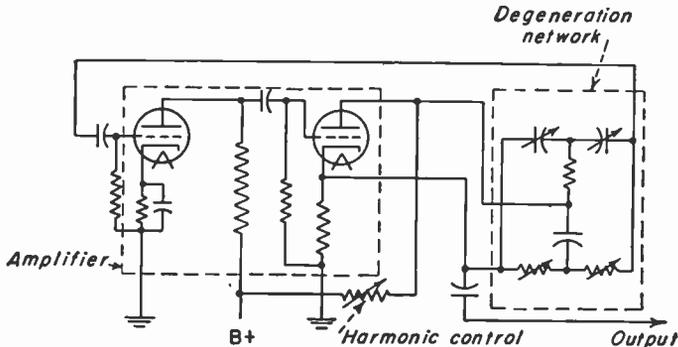


Fig. 17. Resistance-capacitance audio oscillator of Scott.

necessary and "tuning" may be changed by varying resistances. These circuits may be varied over a wide range of frequencies while maintaining a selectivity curve which is a constant percentage function of the "tuned" frequency.

A low-power oscillator operating on the inverse feedback principle has been designed which has exceptionally pure wave form. By the use of an RC network, all frequencies except the frequency of oscillation are fed from the output of an amplifying system back into the input in such a way as to cancel the gain. Regeneration is introduced into the circuit in sufficient amount to cause self-oscillation. This is controlled by the RC network, and hence no inductances or transformers are required in the oscillating circuit.

Figure 18 shows a block diagram which may be helpful in clarifying the action of the system. The circuit includes three separate sections. The section designated A is an amplifier and has substantially flat frequency response and negligible phase shift over the a-f range. The degeneration network, section B, balances to a sharp null at the frequency of oscillation. This provides full amplifier gain at this frequency, and gain at all other frequencies is substantially canceled. The regenerative feedback network, section C, is fed through a phase-reversing tube D to provide

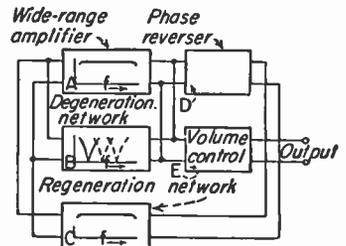


Fig. 18. Block diagram, RC audio oscillator.

¹ Scott, H. H., *Proc. IRE*, 26, 226, 1938; *Gen. Radio Experimenter*, 13, (11), 1939.

the proper regenerative action. Section *C* also has a flat frequency response and is adjusted to provide just sufficient regeneration to produce self-oscillation.

This oscillator covers the frequency range from 20 cycles to 15 kc. Operating under normal conditions approximately 0.25 watt of power may be obtained with less than 1 per cent distortion. With higher outputs the distortion is increased somewhat. This type of oscillator makes possible certain measurements which were previously impractical.

Experimental models of this type of oscillator have been constructed which generate sinusoidal frequencies from 1 cps to 200 kc. By the use of class B operation of the oscillator tubes in push-pull circuits, the power output can be extended appreciably.

16. Negative-resistance Oscillators. In feedback oscillators it can be shown that a necessary condition for the production of sustained oscillation is that the tube, together with the resonant circuit, produce an equivalent negative resistance.¹ As distinguished from the usual feedback oscillators, negative-resistance oscillators are those in which the negative resistance of the system does not require the presence of a tuned circuit.

Oscillators of this type are as follows:

1. Dynatron oscillators
2. Transatron oscillators (negative transconductance)
3. Negative resistance push-pull oscillators
4. Semiconductor circuit element oscillators

17. Dynatron Oscillator. The dynatron oscillator of Hull² (see Fig. 19a) depends for its operation on the phenomenon of secondary emission. He showed that it was possible to use the negative resistance produced by secondary emission for the generation of oscillations. Usually the dynatron oscillator employs a screen grid tube which operates with a plate voltage less than the voltage applied to the screen grid. Under these conditions the characteristic shown in Fig. 19b results. It can be seen that there is an appreciable range in which a positive increment in plate voltage causes a negative increment in plate current, *i.e.*, negative resistance. Secondary emission of electrons at the plate causes this negative resistance characteristic which may be explained as follows: The potentials of the control and screen grids determine largely the number of primary³ electrons which arrive at the plate. The plate potential, however, controls the velocity at which the primary electrons strike the plate. Therefore, the number of secondary electrons⁴ produced at the plate increases as the plate voltage is increased. All the secondary electrons produced are drawn to the more positive screen grid, and the effective plate current is the difference between the primary electrons received at the plate from the cathode and the secondary electrons lost by the plate.

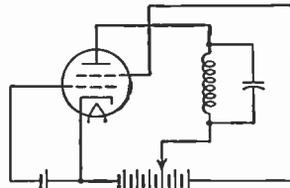


FIG. 19a. Dynatron oscillator.

If, as the plate voltage is increased, more electrons leave the plate owing to secondary emission than arrive from the filament, the effective plate current may decrease. This condition results in a negative dynamic resistance, and the characteristic shown in Fig. 19b is obtained. Oscillation will be developed if a tuned circuit is connected

¹ The a-c resistance of a device may be defined as the reciprocal of the slope of its current-voltage characteristic. If this slope is negative for a certain range in voltage, the device is said to have a negative resistance throughout this range. Under this condition a positive increment in current through the device results in a negative increment of voltage across its terminals. When the direction of flow of d.c. is opposite to the applied direct voltage, as may be observed in certain devices, such devices are said to have a negative d-c resistance.

² Hull, A. W., *Proc. IRE*, 6, 535, 1918.

³ Primary electrons are those which are emitted from the cathode.

⁴ Secondary electrons are those which are obtained from materials as a result of impact of quickly moving electrons which knock electrons out of a solid body when striking with sufficient velocity. One primary electron striking a material at high velocity may produce many secondary electrons.

across this negative resistance as shown in Fig. 19a, provided the absolute value of the negative resistance is less than, or equal to, the equivalent resistance of the tuned circuit. The amplitude of oscillation may be varied by means of the control-grid voltage, which varies the slope of the current-voltage characteristic in the negative-resistance range.

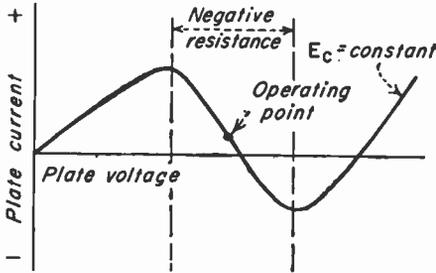


FIG. 19b. I_p - E_p curve, dynatron oscillator.

When designing a dynatron oscillator, the point of operation should be chosen to be in the center of the most linear region of the negative-resistance characteristic,¹ and the amplitude of oscillation should be kept small. Under these conditions the curvature in the operating range can be kept small and the harmonic content low.

In addition to excellent wave form, the dynatron oscillator possesses good frequency stability and simplicity.

The chief disadvantage to this type of oscillator arises from its dependence upon secondary emission, a property which is extremely variable with age and which varies widely in tubes of the same type. With tubes of ordinary size the power output is extremely limited.

18. Transitron Oscillator (Negative Transconductance Oscillator).² The name *transitron* has been proposed by Brunetti for the retarding-field negative-transconductance oscillator. This oscillator possesses essentially the same type of negative-resistance characteristic as the dynatron oscillator and has all its advantages without its disadvantages. Its characteristic is independent of secondary emission and remains practically constant throughout the life of the tube. The action of this oscillator, shown in Fig. 20a, is as follows:

The suppressor voltage is chosen so as to make the suppressor grid negative with respect to the cathode. Electrons that have passed through the screen grid are repelled by the suppressor grid and return to the screen because of its high positive voltage. Hence the suppressor grid with its retarding field acts as a virtual cathode. A small negative increment in voltage across the tuned circuit is transmitted to both the screen and suppressor grids, causing the suppressor grid to repel more electrons and the current to the screen grid to increase. Hence the transconductance between the screen and suppressor grids is negative. The characteristic current-voltage curve for this type of oscillator is as shown in Fig. 20b.

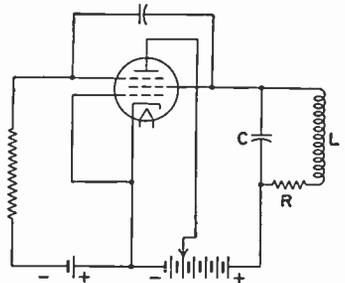


FIG. 20a. Transitron oscillator.

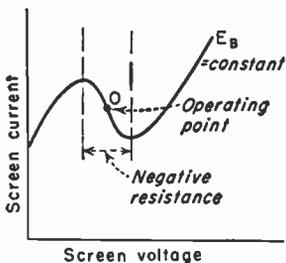


FIG. 20b. Screen characteristic, transitron oscillator.

At the operating point O , oscillation in the resonant circuit will begin. If L/RC is increased, the

¹ See Terman, F. E., "Measurements in Radio Engineering," p. 289, McGraw-Hill, 1935.

² Herold, E. W., *Proc. IRE*, **23**, 1201, 1935. For an excellent treatment on the practical design of the transitron oscillator see C. Brunetti, *Proc. IRE*, **27**, 88-94, 1939.

amplitude of oscillation increases. As with the dynatron oscillator, it is desirable to keep the amplitude of oscillation small so as to keep the wave form and frequency stability good.

When a small negative bias is applied to the control grid, the total current flowing to the screen grid may be controlled, and the negative slope of the current-voltage characteristic may be varied. Hence a flexible means is available for varying the magnitude of the negative resistance and thus the amplitude of oscillation. By having the oscillation voltage regulate the bias on the control grid, additional amplitude control may be obtained.

Like the dynatron oscillator, this is essentially a low-power oscillator. It will generate sinusoidal oscillations of any frequency from the lower audio to approximately 60 Mc by simply changing the tuned circuit constants. Suitable pentodes for the transitron oscillator are the types 57, 58, 59, 89, 6C6, 6J7, and 6K7. In a properly designed oscillator, Brunetti reports that changes resulting from a 33 per cent change in direct screen-grid voltage may be kept within 10 parts in 10^6 and that, in general, the transitron oscillator frequency stability may be compared with that of a crystal oscillator without temperature control.

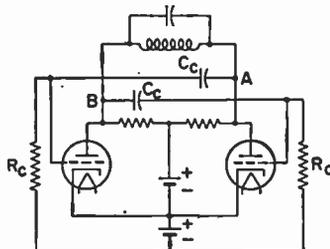


FIG. 20c. Push-pull negative resistance oscillator.

19. Push-pull Negative-resistance Oscillator. A negative-resistance oscillator of low harmonic content and excellent frequency stability can be designed employing two tubes in push-pull as shown in Fig. 20c.¹ The action of this circuit is as follows. If the two tubes have identical characteristics and if the voltage between *A* and *B* is zero, the two plate currents are equal, and there is no current flowing through *A* and *B*. When an increment of voltage is applied between *A* and *B*, an increment of current will flow which will raise the plate voltage and lower the grid voltage of one of the tubes and lower the plate voltage and increase the grid voltage of the other tube. When this voltage is sufficiently small, the plate resistance and transconductance are substantially constant. If the amplification is large enough, the change in plate current exceeds the current flowing between *A* and *B* and is opposite in direction to the applied voltage. This results in a current flowing through the network between *A* and *B* which is opposite in direction to the applied voltage, and a negative resistance is obtained. When a parallel resonant circuit of high selectivity is connected between these terminals, sustained oscillations are developed.

The amplitude of oscillation may be readily controlled by means of the grid bias. When the reactance of *C* is small in comparison with R_c at the lowest desired frequency, which condition it is necessary and desirable to meet, the resonant circuit can be connected between either the two plates or the two grids.

20. Semiconductor Circuit Element Oscillators. Some semiconductors, such as germanium,² silicon³ and other crystals, and thermal resistors,⁴ under certain conditions, exhibit negative resistance characteristics which suggests their use in oscillator circuits. The slope of the voltage-current curve of a germanium crystal diode, shown in Fig. 21, illustrates the negative resistance feature of crystal diodes. Basically, the crystal diode consists of a small-diameter pointed "whisker" of metal, usually tungsten, making small area contact with a semiconductor. The crystal diode is generally used as an u-h-f rectifier, with operation limited to the region near the origin where the resistance is extremely high to current flow from semiconductor to metal. However, in following the characteristic curve in the negative voltage region, it is seen

¹ Reich, H. J., *Proc. IRE*, **25**, 1387, 1939; also Turner, L. B., *Radio Rev.*, **1**, 317, 1920.

² Cornelius, E. C., *Electronics*, February, 1946.

³ Stephens, W. E., *Electronics*, June, 1946.

⁴ Thermal resistors are circuit elements having large negative-temperature coefficients of resistance. For a discussion of thermal resistors, see G. L. Pearson, *Bell Lab. Record*, **19** (IV), December, 1940.

that the dynamic resistance becomes zero and then negative, as the current becomes larger. The voltage developed across the diode then decreases as current increases. This is a temperature dependent action, in that the heating caused by the increased current is responsible for the shape of the voltage-current curve.

A sine-wave oscillator circuit,¹ making use of the high negative dynamic resistance

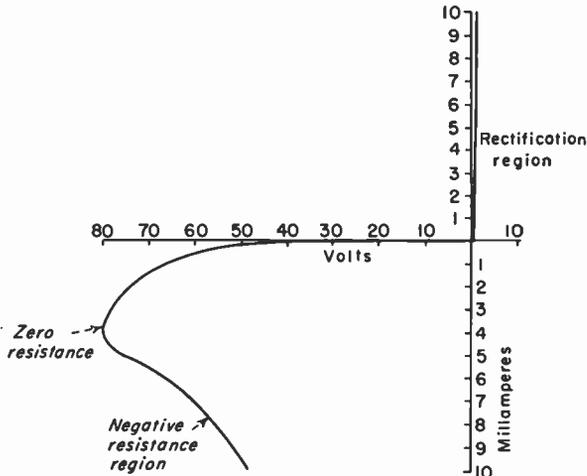


FIG. 21. Voltage-current curve, germanium crystal diode.

features of the crystal diode, is shown in Fig. 22a. A series-resonant circuit is used to control oscillations, since its low impedance at resonance allows sufficient current for negative-resistance operation of the crystal diode. Oscillations up to 1 Mc may be obtained.

A relaxation oscillator circuit,¹ making use of zero resistance characteristic of the crystal diode to provide rapid discharge of the capacitor, is shown in Fig. 22b. Oscillation frequencies up to 500 kc are possible.

Thermal resistors exhibit somewhat the same temperature-dependent

type of negative resistance, but differ from crystal diodes in that heat is supplied, not from passage of current through the resistor, but from an external source. Because of the thermal capacity of the thermal resistors, oscillations that may be obtained are limited to the audio range. Pearson² describes such an oscillator which has been in continuous operation for over 2 years without serious signs of deterioration of the thermal resistor.

In addition to the above types of nonlinear circuit elements, directly heated thermistors, such as are commonly used for microwave power measurement, exhibit negative-resistance characteristics which, in turn, suggest their use in oscillator circuits.

21. Low-frequency Oscillators. At very low frequencies standard circuits become impractical. Capacitors, and particularly inductances, required become very bulky and expensive. Accordingly, certain rather special methods of obtaining low frequencies have been resorted to. The heterodyne and RC oscillators are excellent for this purpose. RC oscillators are described in Sec. 15.

22. Beat-frequency Oscillators.³ By beating together (heterodyning) two r-f voltages of slightly different frequencies, a-f energy may be generated. Oscillators

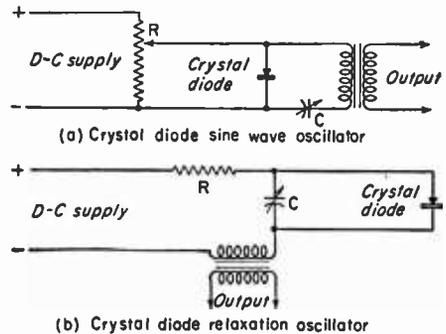


FIG. 22. Oscillator circuits employing crystal diodes.

¹ Cornelius, *loc. cit.*

² Bell Lab. Record, *loc. cit.*

³ An excellent discussion of beat-frequency oscillators is given by F. E. Terman, "Measurements in Radio Engineering," p. 298, McGraw-Hill, 1935.

operating on this principle are called *beat-frequency* or *heterodyne* oscillators. A block diagram of such an oscillator is shown in Fig. 23. The outputs of two r-f oscillators of slightly different frequencies are applied at the same time to a detector. In addition to the impressed frequencies, the output of the detector contains their sum and difference frequencies. The filter, shown connected to the output of the detector, removes the fundamental radio frequencies and their sum and leaves only the difference frequency which may be amplified as desired.

Among the advantages of this type of a-f oscillator is the fact that the whole range of audio frequencies may be obtained by tuning a single dial. Another advantage is that the use of large coils and capacitors such as are employed in other types of a-f oscillators is avoided, resulting in lightness and compactness.

A number of special problems arise in the design and construction of beat-frequency oscillators. One of these problems is to eliminate the tendency of the two oscillators

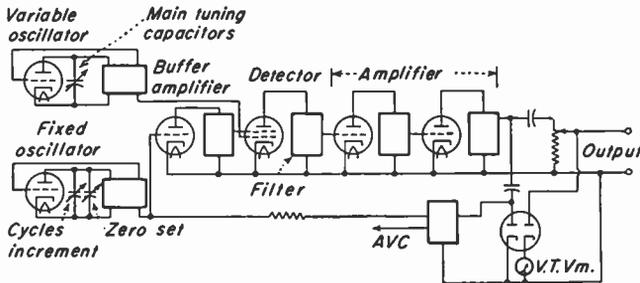


FIG. 23. Block diagram beat-frequency oscillator.

to pull into synchronism when their frequency difference is small, *i.e.*, when low audio frequencies are being produced.

This tendency to interact may be avoided by proper shielding, careful arrangement of the component parts, proper use of decoupling resistors, choke coils, and by-pass capacitors, and by the use of special methods of coupling the oscillators to the detector. The most frequently used methods of coupling are (1) the use of a buffer amplifier between each oscillator and the detector, (2) the use of a balanced modulator circuit, or (3) the use of electron-coupled oscillators.

To avoid beats between harmonics generated by the r-f oscillators, an r-f filter is placed between the r-f oscillators and the detector. The fixed oscillator should have a smaller voltage output than that required by the variable-frequency oscillator in order that distortion of the output be reduced.

Higher order curvature of the detector characteristic produces additional distortion of the output, which may be prevented by the use of the balanced modulator. When square-law detectors are used, this type of distortion may be reduced by correct adjustment of bias and input voltage. Distortion produced by linear detectors may be reduced by making the output amplitude of one of the r-f oscillators small in comparison with that of the other.

The frequency stability of the output of beat-frequency oscillators is generally poor. This is because a very small percentage variation in frequency in the output of one of the r-f oscillators will result in a comparatively large percentage variation in the a-f output.

By making the two r-f oscillators as nearly identical as possible, they may be made to react similarly to variations in temperature, polarizing potential, etc., and thereby the effects of these quantities may be minimized. The h-f oscillators employed are usually stabilized by one of the methods discussed in Sec. 3 or by the use of negative-resistance oscillators. To compensate frequency drift in beat-frequency oscillators, a small trimming capacitor is always provided which can be adjusted so that a particular

point on the frequency calibration is correct. This point is obtained either by comparison of the output frequency with a standard frequency source or by using the zero-frequency point.

In the output circuit of the detector, it is desirable to install a low-pass filter. This filter prevents the overloading of the a-f amplifier due to r-f voltages that may exist in the detector, and hence improves the output wave form.

The frequency at which the r-f oscillators operate is usually between 100 and 500 kc. At these higher frequencies the differences between the design constants of the fixed- and variable-frequency oscillators are less. This allows more nearly identical design, which, as pointed out above, leads to better frequency stability for the a-f output. Also, the filter requirements are simplified by the use of the higher frequencies. On the other hand, the a-f stability is decreased as the r.f. is increased, and commercial design usually fixes 500 kc as the upper limit. The General Radio Company has produced an excellent heterodyne oscillator extending to 5 Mc. The fixed frequency in this range is 20 Mc.¹ Western Electric type 17A and 17B oscillators cover the frequency band from 0 to 150 kc and are accurate to ± 50 cycles throughout the band. Correct settings at 60 cycles and 100 kc are obtained by special built-in apparatus and trimmers.

23. Relaxation Oscillators.² Relaxation oscillators are used for the generation of distinctly nonsinusoidal waves. In certain applications this type of oscillator

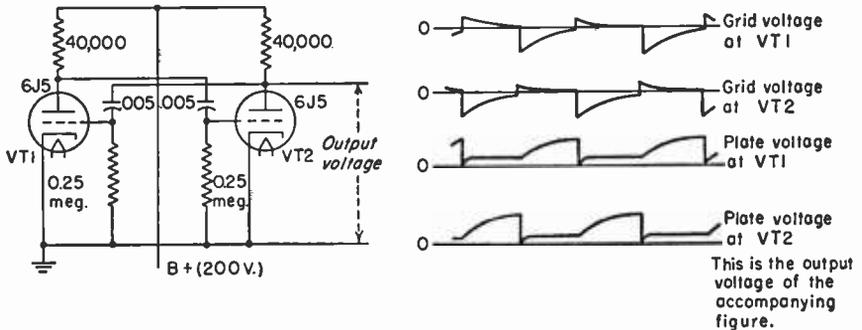


FIG. 24. Fundamental circuit of the multivibrator and the voltage relations of the various branches.

has many advantages over the resonant circuit oscillators. Its output is extremely rich in harmonics, but its frequency, which is not very definitely fixed by the circuit elements, may be easily stabilized by the introduction of small voltages of harmonic or subharmonic frequency into the oscillating system. Relaxation oscillators are also comparatively inexpensive, simple, and compact and can be conveniently designed to cover a wide range of frequency.

The process by which relaxation oscillations are produced involves the building up and breaking down of the energy stored in the electric field of a capacitor or the magnetic field of an inductance. Various devices may be used to control this building-up and breaking-down action, such as glow or arc-discharge tubes or high-vacuum tubes.

Among the relaxation oscillators employing high-vacuum tubes is the multivibrator.¹ The multivibrator, which is most satisfactory for frequency conversion, was the first relaxation oscillator to be developed. Figure 24 shows the basic circuit. The voltage drop across any of the circuit elements may be taken as the output voltage, and the frequency of oscillation, which is a function of every circuit parameter, may be controlled by variation of the resistances and capacitances and is approximately equal to

¹ *Gen. Radio Experimenter*, January, 1939.

² Abraham, H., and E. Block, *Ann. Physik*, 12, 237, 1919.

$f = 1/(r_1C_1 + r_2C_2)$. When the circuit is symmetrical, the wave form of the grid and plate voltages is as shown in Fig. 24.

An unsymmetrical multivibrator may produce an output wave in which the positive and negative periods differ by an appreciable amount. This is accomplished by making the time constant of one of the grid circuits short compared to the time constant of the other grid circuit.

The multivibrator may be used for generation of submultiple frequencies by injection of an alternating control voltage of frequency equal to or higher than the natural frequency of the multivibrator, in the multivibrator circuit, the resulting oscillations being an exact submultiple of the control frequency. Control of the submultiple frequency depends on the magnitude of the control voltage, the manner in which the control voltage is applied to the grids, and the degree of symmetry of the multivibrator circuit. If, in a symmetrical circuit, the control voltage is injected so as to affect each grid in the same manner, the resultant output will favor even submultiples of the control frequency; likewise, under the same conditions, if the effect of the control voltage is opposite in phase at each grid, the output will favor odd submultiples.

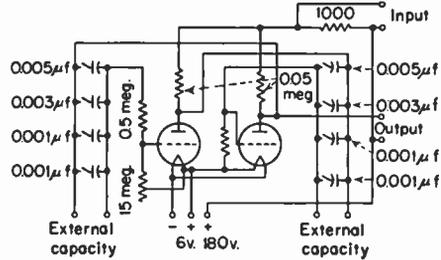


FIG. 25. Generator of submultiple frequencies.

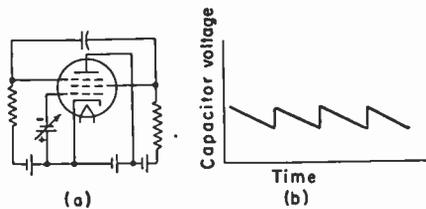


FIG. 26. (a) Van der Pol relaxation oscillator using pentode. (b) Typical wave form of capacitor voltage.

Submultiple frequencies as low as one-fourteenth of the input frequency can easily be had. External capacitance, as indicated, must be added at the l-f range. By appropriate choice of the circuit constants, the fundamental frequency range may be extended to 10^{-1} cps.

Another form of relaxation oscillator employing a high-vacuum tube was originally described by van der Pol.¹ The circuit for this oscillator is shown in Fig. 26a, and the wave form of the capacitor voltage is shown in Fig. 26b. This type of wave form, which is known as a *saw-toothed voltage* wave, is used in connection with cathode-ray oscilloscopes and cathode-ray television tubes. The circuit of Fig. 26 is identical to the transitron circuit of Fig. 20 except for the use of a tank circuit.

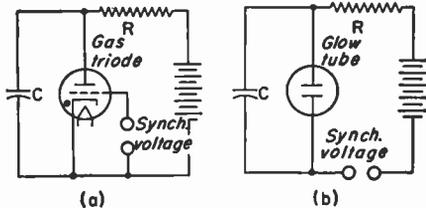


FIG. 27. Relaxation oscillators using gas-filled tubes.

Relaxation oscillators for generating saw-toothed wave forms are often designed using grid-controlled gas-filled triodes.² A property of these tubes that makes them

¹ van der Pol, B., *Phil. Mag.*, 2, 978, 1926.

² For an excellent discussion and design data for relaxation oscillators using gas-filled tubes see "Measurements in Radio Engineering" by Terman, pp. 315-322, McGraw-Hill, 1935.

suitable for this purpose is their so-called *trigger* action. If their grid potential is momentarily less than the cutoff value, positive ions are produced in the tube which neutralize the negative space charge of the electrons, as well as the controlling action of the grid. This results in a very rapid change in the plate resistance of the tube from a high value to a very low value. The time required to ionize and deionize the gas in the tube limits the frequency for which oscillators of this type can be built.¹

A basic circuit for a relaxation oscillator using a gas-filled tube is shown in Fig. 27a. The action of this circuit is as follows: The direct plate voltage charges *C* through *R* until the critical starting potential of the tube is reached. At this potential the positive ions are produced, and the resistance of the tube falls to a very low value, discharging *C*. When the plate voltage drops to a certain value, the plate resistance returns to its original high value, and the cycle is repeated. The value of the grid polarizing potential controls the critical plate potential at which ionization takes place.

Small alternating voltages may be introduced into the grid circuit for synchronizing purposes as shown in Fig. 27a. If a glow tube, *i.e.*, neon tube, relaxation oscillator is used, the synchronizing voltage may be introduced as shown in Fig. 27b.

A complete circuit diagram of a system suitable for producing saw-toothed wave forms for a cathode-ray oscilloscope is shown in Fig. 28. In this circuit a pentode is used to maintain a constant charging current. By varying its grid bias, the magnitude of the charging current may be controlled.

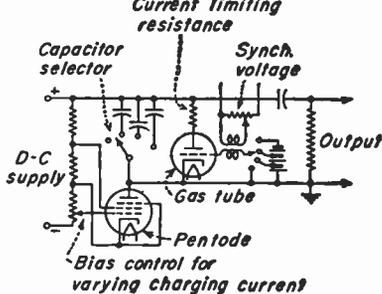


FIG. 28. Relaxation oscillator using pentode to maintain constant driving current.

modulation of the transmitted r-f output.

Typical circuits for production of pulses are as follows: limiters or clippers, over-driven amplifiers, differentiator amplifiers, pulse-shaping lines. Combinations of these circuits follow the conventional oscillator, and each successively distorts its input wave form to ultimately produce the pulses. The stability of the pulse repetition frequency depends on the stability of the oscillator, which thus performs the basic timing function.

25. Limiters. A limiter removes part or all of one or both polarities of an input wave. Either diodes or triodes, tetrodes, and pentodes may be used in the limiting circuit, resulting in diode or grid limiting, respectively. In diode limiting, the rectifying ability of the diode causes it to act as a switch. Various circuit arrangements are shown in Fig. 29. When the input wave drives the diode plate positive, the diode conducts and passes that portion of the wave. Similarly, when the plate goes negative, the diode blocks passage of the wave. A partial diode limiter may be so connected and biased that only a portion of the negative or positive peak is removed.

¹ Oscillators of this type have been built to operate successfully as high as 20,000 cps.

² This concept may be applied to the a-c component of the output curve of a relaxation oscillator operating in the usual manner, *i.e.*, on the linear portion of the capacitor charge curve.

24. Impulse Generators. By the use of special circuits, wave forms known as "pulses" may be produced from the output of conventional oscillators. The wave forms are dissimilar to those associated with sine-wave, symmetrical multivibrator, or relaxation oscillators, in that the recurrence of the pulses is not restricted to equal partial periods.² The production of the pulses depends on successive distortion of the output wave forms of conventional oscillators. Typical uses of pulses are in television systems where short-duration rectangular pulses synchronize line and frame sequence and in radar systems where pulses perform various timing and trigger functions, including

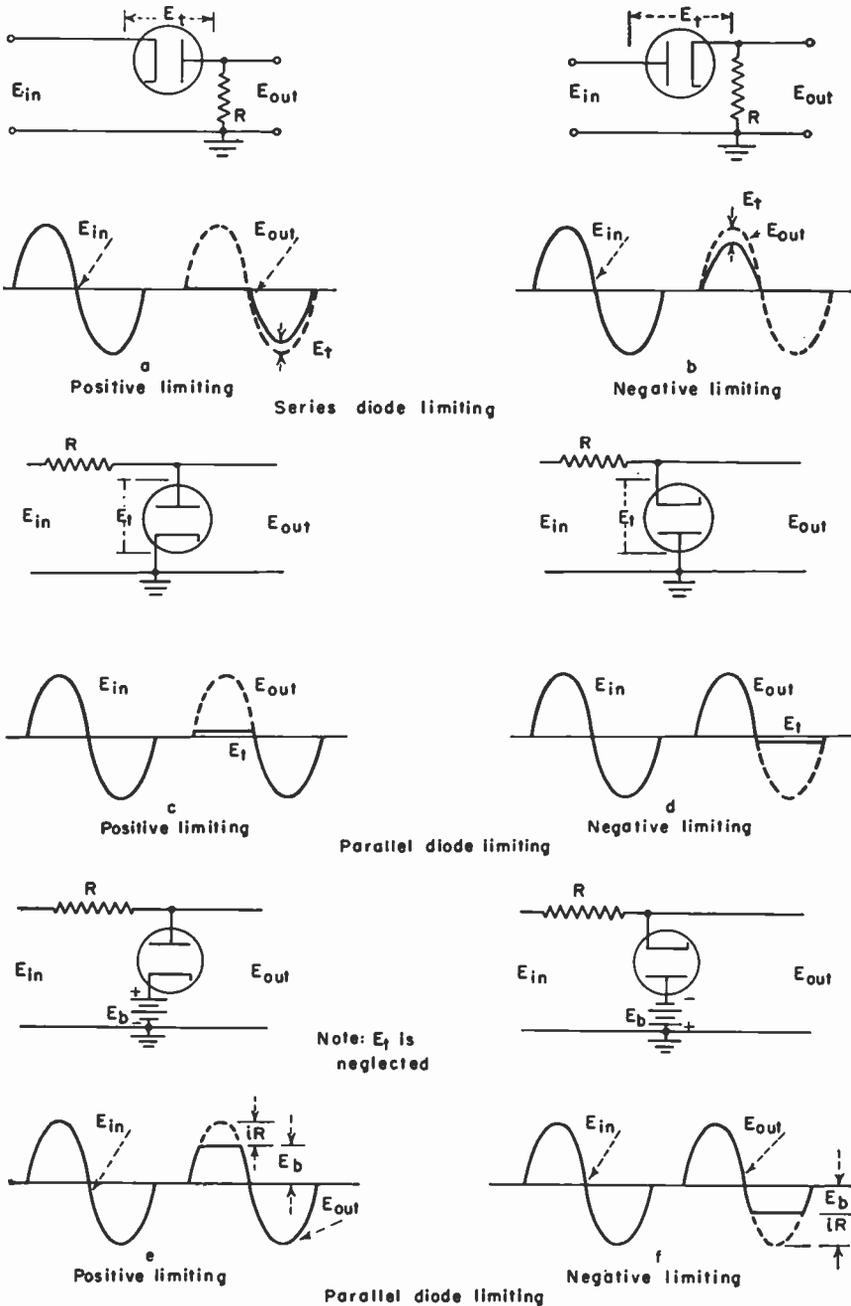


FIG. 29. Diode limiter circuits.

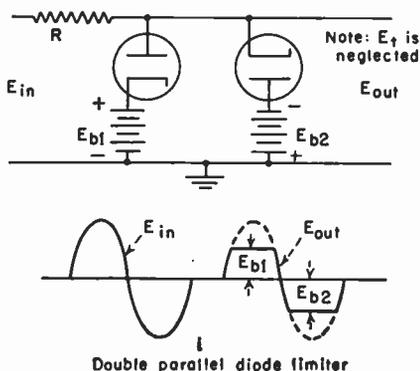
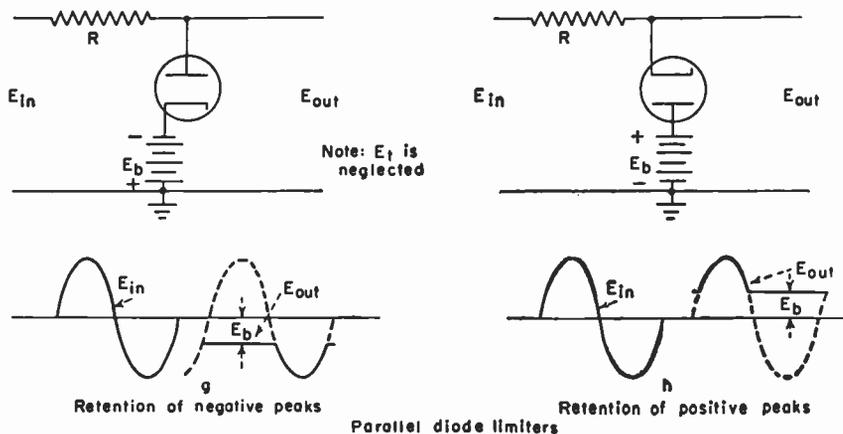


FIG. 29. (Continued.)

In grid limiting, the limiting action is performed in the grid-cathode circuit. A typical circuit using a triode is shown in Fig. 30. When the input wave swings positive, the grid attempts to swing positive, resulting in an enormous decrease in grid-to-cathode resistance. However, the voltage drop developed across the resistor, R , opposes the positive input voltage and results in cancellation except for the small voltage across the grid-to-cathode path. As in the case of the parallel diode limiter, the grid limiter may be biased to remove only the top portion of the positive half sinusoid.

approximately square waves from sine waves. A circuit similar to the grid limiter, but without the limiting resistor, is used to obtain plate-current saturation on the positive peaks of the input signal and plate-current cutoff on the negative peaks. The circuit diagram of an overdriven amplifier is shown in Fig. 31a, and typical operating curves are shown in Fig. 31b. It may be seen from Fig. 31b, that for a relatively low plate-supply voltage and high load resistance, a small positive input signal will cause plate saturation. For limiting action on the negative peak, a high- μ tube is used because of its low cutoff characteristics. The effect of saturation and cutoff is the desired "squaring" of the plate current. If the input signal contains only positive or negative half sinusoids, owing to a previous limiting action, the plate current will contain only positive or negative portions of square waves.

27. Differentiator Amplifiers. The differentiator amplifier consists of short-time constant RC circuit followed by an overdriven amplifier normally operating at near saturation plate current. Positive peaks of voltage applied to the grid do not affect

26. Overdriven Amplifiers. An overdriven amplifier is used to produce ap-

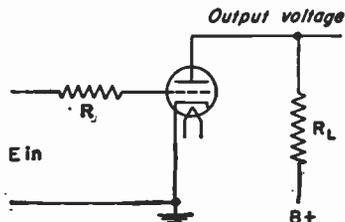


FIG. 30. Unbiased grid limiter circuit.

the plate current and hence the plate voltage; negative peaks applied to the grid cause the plate current to decrease, thus allowing a sharp rise in plate voltage. When the leading edge of a steep wave front is applied to the RC circuit, the current immediately assumes the value E/R , then decreases as the capacitor charges. (E = peak voltage of wave front, R = resistance of RC circuit.) For a square-wave input, the current decrease is exponential. The voltage across the resistor and the amplified output are shown in Fig. 32. Successive stages of differentiation serve to produce wave forms with higher peaks and shorter time bases.

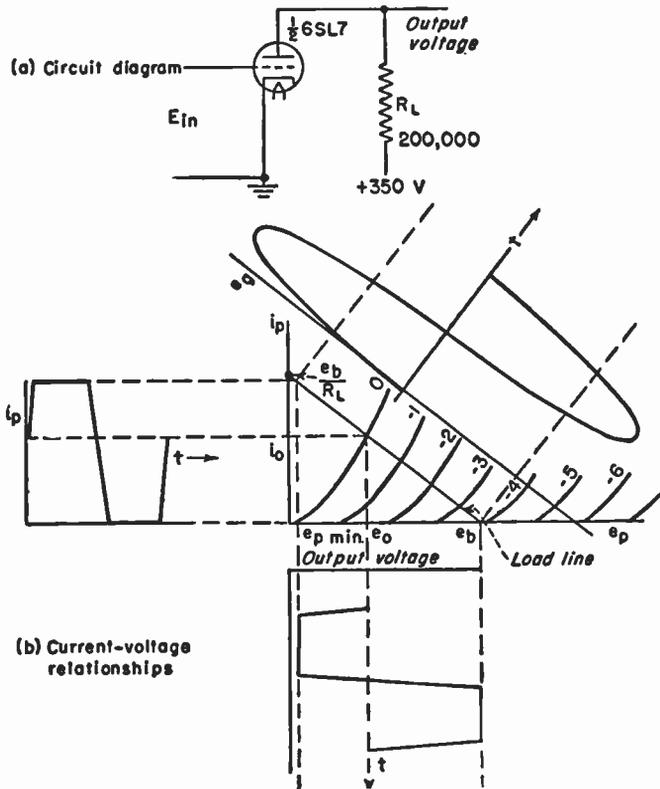
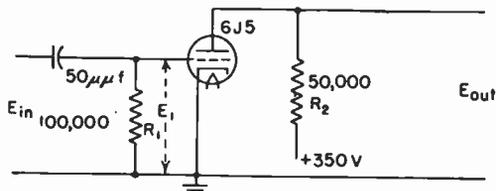


FIG. 31. Overdriven amplifier circuit.

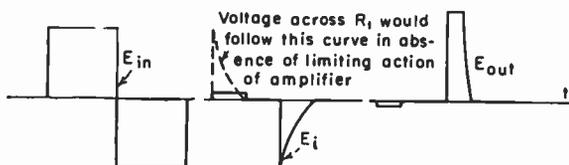
28. Pulse-shaping Lines. An artificial transmission line consisting of series L and shunt C terminated in a short circuit is used as an energy-storing device for the production of rectangular pulses. When a high potential is suddenly applied to the line, the resultant wave form at the terminals is a rectangular wave of duration equal to the time required for the applied wave front to travel to the end of the line and be reflected to the source. The time duration of the rectangular wave is precisely determined by the constants of the line and is not dependent on the applied potential. An electronic switch is used to prevent the pulse from being reflected back into the source.

29. Complete Pulse-forming Circuits. The three most commonly used vacuum-tube timing oscillators for pulse forming are as follows: sine-wave timing oscillators, multivibrator timing oscillators, blocking oscillators.

30. Sine-wave Timing Oscillators. The sine-wave timing oscillator method employs a sine-wave generator as the basic oscillator. The block diagram of Fig. 33



(a) Circuit diagram



(b) Voltage relations

FIG. 32. Differentiator amplifier.

charge on the grid capacitor drives the grid beyond cutoff after one complete cycle. After a time determined by the RC constant of the grid resistor and capacitor, the

shows the sequence of circuit elements and, qualitatively, the wave forms produced at each stage.

31. Multivibrator Timing Oscillators. An asymmetrical multivibrator is often used as the basic oscillator in a pulse-forming circuit. The block diagram is shown for this circuit in Fig. 34.

32. Blocking Oscillator. The circuit diagram of a blocking oscillator is shown in Fig. 35. This relaxation oscillator forms the basic generator for the blocking oscillator timing circuit. If the grid resistor is sufficiently large, the accumulation of

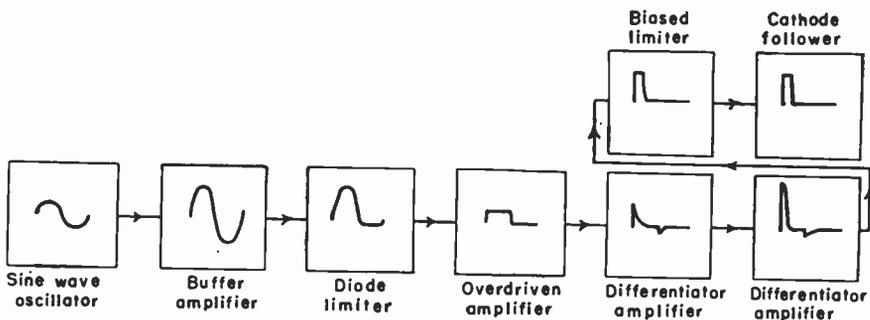


FIG. 33. Sine-wave timing oscillator.

operation is repeated. One of the half sinusoids is next removed by the limiter. The output of the limiter is applied to a pulse-shaping line as shown in Fig. 36. A biased

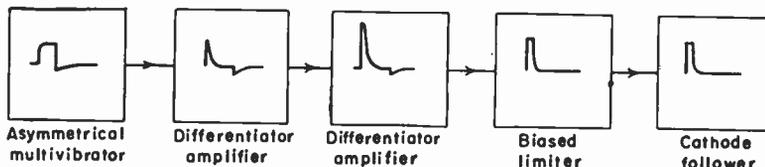


FIG. 34. Multivibrator timing oscillator.

limiter removes part of the top of the output wave from the pulse-shaping line to produce the short-duration rectangular pulse.

33. High-frequency Oscillators. Generation of h-f power has become extremely important, with respect to both commercial and military applications. Although practically all present-day tubes may be depended on to operate satisfactorily at frequencies as high as 30 Mc (some as high as 50 Mc with full ratings and 70 Mc with reduced ratings), conventional-type tubes especially designed for h-f use and tubes such as the magnetron and klystron have extended the frequency range considerably.

As the frequency range is increased, resonant lines and cavities replace the lumped circuit constants as control elements. The selectivities of resonant lines and cavities (Q ranging from 1,000 to 50,000) compare favorably with crystals. It should be noted that, while cavities and resonant lines have high Q 's, the stability of these structures is not necessarily good over long periods of time or when changes in temperature are encountered, inasmuch as expansion and contraction cause changes in the physical dimensions of the structures.

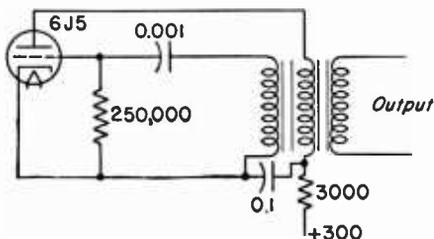


FIG. 35. Blocking oscillator.

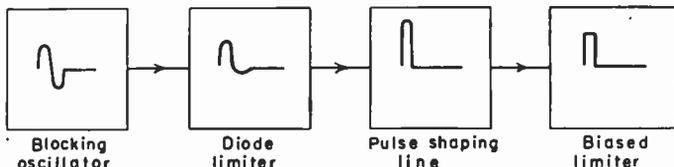


FIG. 36. Blocking oscillator timing circuit.

When conventional triodes are used in h-f oscillators, a top frequency limit is reached due to interelectrode capacitances, lead inductances, and finite transit time of the electrons.

A h-f triode oscillator, known as the "ultraudion" oscillator, is illustrated in Fig. 37.

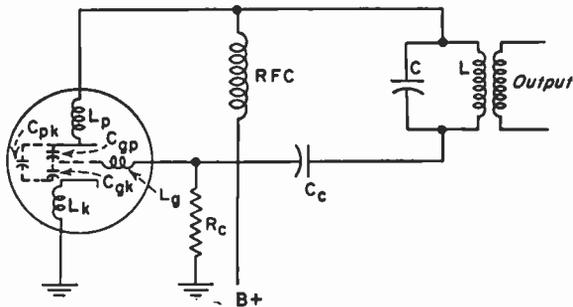


FIG. 37. Ultraudion oscillator showing lead inductance and interelectrode capacitances.

This oscillator may be considered as a modification of the Colpitts oscillator, wherein use is made of capacitances C_{gp} and C_{gk} to perform the voltage division between the plate and grid circuits.

To increase the frequency range, the tank circuit, LC , may be reduced in size until it appears as a short section of wire. It is seen that the lead inductances and interelectrode capacitances

eventually determine the limiting frequency at which this particular oscillator will operate. Moreover, since L_k is an impedance common to both the plate and grid circuits, degenerative feedback voltage developed across this impedance will reduce the amplitude of oscillation further.

When the time required for electrons to travel from cathode to plate is not negligible compared to the time for one cycle of oscillation, power is consumed by the grid circuit

acting as a resistance which varies inversely as the square of the frequency.¹ This fact is another reason why the amplification falls off to a point where oscillations will no longer be sustained when the frequency is increased sufficiently.

To increase the frequency range, tubes have been designed with low interelectrode capacitances, lead inductances, and with elements closely spaced to reduce transit-time

effects. The so-called "acorn" and "doorknob" tubes are representatives of this type, with top frequencies in the order of 1,500 kc. Small elements closely spaced, however, limit the power-handling ability of these h-f tubes.

An oscillator circuit which employs a concentric line as a frequency-control element is shown in Fig. 38. This is a standard tuned-plate tuned-grid feed-back oscillator in which the grid resonant circuit is replaced by the concentric line. For best stability, the line should be loosely coupled to the grid circuit. This is accomplished by making the grid connection at a point comparatively close to the shorted end of the line. By proper

FIG. 38. Oscillator employing a resonant line as frequency-control element.

adjustment of the regeneration control C_1 , the phase of the grid excitation may be advanced so as to compensate for the phase lag of the electron current in the tube. The length of the connections between the tube and the resonant circuits must be small compared with the wave length at which the system is operating. This condition may be very difficult to meet at extremely high frequencies, and special circuits have been designed which help avoid this difficulty.

Peterson² uses a high- Q (approximately 2,000 in frequency range of 60 to 140 Mc) resonant circuit, consisting of an outer cylinder with a cylindrical shaped insert, as the tank circuit in a modified, tuned-plate oscillator (see Fig. 1b). Frequency stabilities of the order of 5 parts in 10^6 for a 50 per cent change in plate voltage have been obtained. One of the chief advantages, in addition to its excellent stability, is that difficulties arising from tube connections are greatly minimized (see Fig. 39). Also, the size of the resonant circuit is only a fraction of its equivalent concentric line. When a continuously variable oscillator is required, however, other arrangements must be used, for only slight variations in frequency can be obtained with this design.

A continuously adjustable stabilized oscillator has been designed by Barrow³ for the frequency range from 70 to 700 Mc. The oscillator circuit consists of a coaxial line that is easily and rapidly adjustable over the entire frequency range. Among

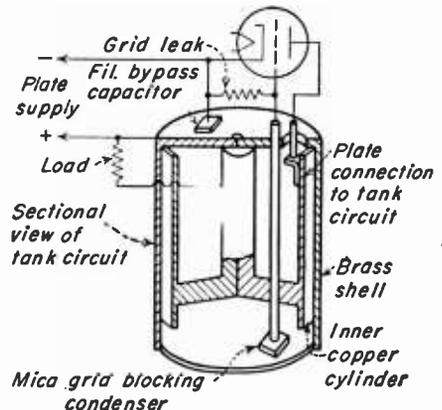
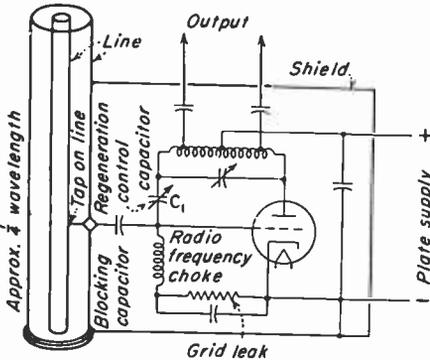


FIG. 39. U-h-f oscillator of Peterson.

¹ North, D. O., *Proc. IRE*, 24, 108, January, 1936.

² Peterson, *Gen. Radio Experimenter* 12, October, 1937; *Communications*, 17, 26-28, 1937.

³ Barrow, W. L., *Rev. Sci. Instruments*, 9, 170, 1938.

other things, it affords excellent shielding, mechanical ruggedness, and a coaxial line output connection. Several watts output are obtained over the entire frequency range. Both filament leads are tuned in addition to the tank. The connections are shown in Fig. 40. At frequencies below 300 Mc the stability is roughly 100 parts in

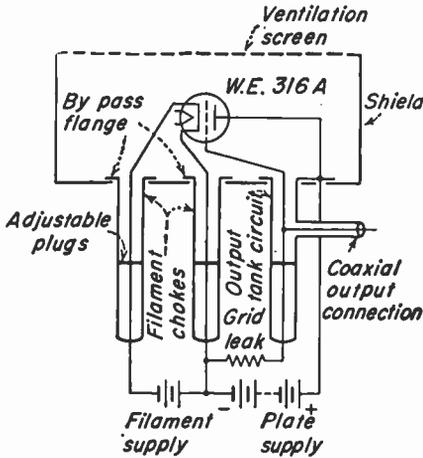


FIG. 40. U-h-f oscillator of Barrow.

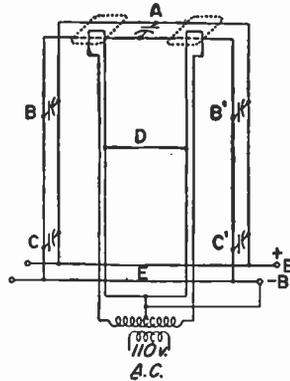


FIG. 41. U-h-f oscillator of King.

10^4 and decreases with increased frequency, becoming very poor near the limit of oscillation of the tube.

Special triodes in which the plate and grid leads provide support for the electrodes and extend through the bulb are especially useful for u-h-f work. A tube of this type having a rated output of 1.5 watts at 1,500 Mc is available commercially. A pair of such tubes has been built into a special oscillator, having a continuous range. This oscillator, designed by King,¹ is particularly suitable for parallel line measurements at ultrahigh frequencies (see Fig. 41). It consists essentially of a rectangle of parallel conductors which may be bridged by blocking capacitors. The frequency depends on the dimensions of the circuit *BAB'* in the figure. For the highest frequencies the capacitor *A* must be used, and capacitors *B* and *B'* must be moved up as close to the triodes as possible. For lower frequencies *A* need not be used. The oscillator is coupled to parallel lines by placing it below the lines.

Circuits employing parallel lines are often used for the generation of u-h-f waves. Representative circuits are shown in Fig. 42. Although these circuits are comparatively simple to construct, the tube is connected at a high impedance point and the frequency stability is poor.

The chokes that appear in the filament leads of the circuits of Fig. 42 are made

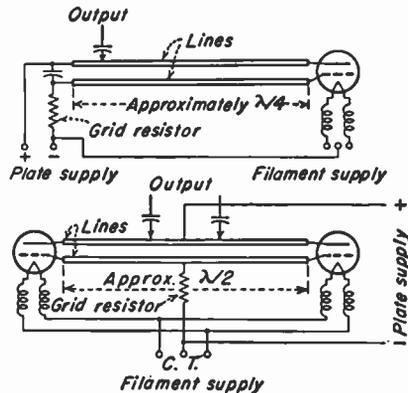


FIG. 42. U-h-f oscillators using parallel lines as circuit elements.

¹ King, R., *Rev. Sci. Instruments*, 11, 270, 1940.

necessary for u-h-f operation, because the filament with its leads may often be a considerable portion of a wavelength and thereby prevent normal operation of the oscillator. A method which is preferred to the use of choke coils is the provision for tuning the filament-to-ground circuit. The use of adjustable concentric lines of approxi-

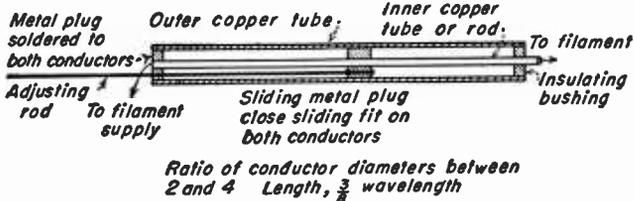


FIG. 43. Adjustable concentric line for use as filament choke.

mately one-fourth wavelength for each filament lead is probably the most satisfactory method (see Fig. 43). It is desirable at u-h-f operation to avoid the use of dielectric material as much as possible and to confine that which is necessary for mounting the circuit elements to points of low r-f potential.

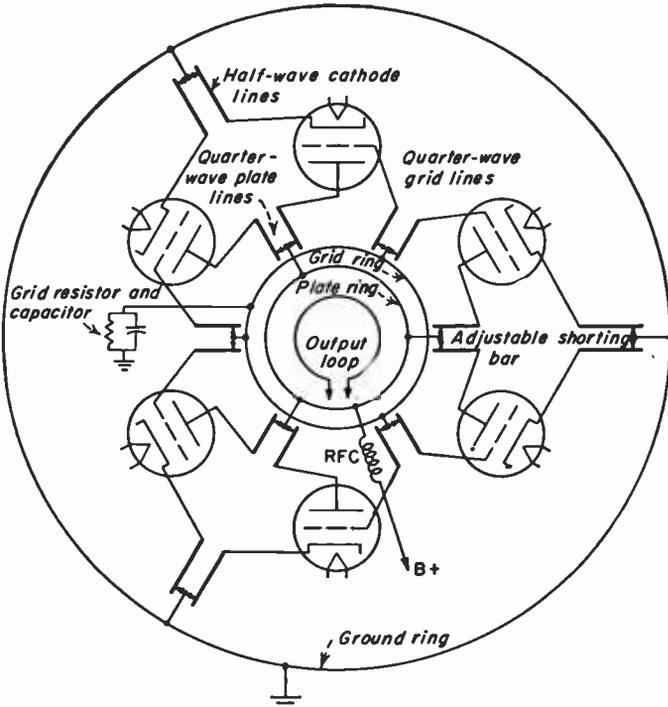


FIG. 44. Six-tube ring oscillator schematic.

With the ring oscillator developed during the recent war, considerable power output at frequencies higher than 300 Mc may be obtained. The circuit is developed from the basic push-pull tuned-plate, tuned-grid oscillator with shorted quarter-wave resonant lines replacing the conventional plate and grid-tank circuits. Four or more similar u-h-f triodes may be arranged in a circle and connected as shown in Fig. 44.

The cathodes are connected to shorted half-wave lines, the shorting bars of which are grounded by means of the ground ring. This effectively places the cathodes at r-ground potential. Plate power is supplied through the plate ring which is connected to the shorting bars of the plate resonant lines. The tubes are biased by means of the grid-leak and combination capacitor which is connected to the grid ring and thence to shorting bars of the grid-resonant lines. Output power is obtained through magnetic transfer of energy from the shorting bars of the plate lines to the output coupling loop. Oscillations are sustained due to feedback through each plate-grid capacitance and inherent magnetic coupling between the plate and grid tuned circuits.

High-frequency oscillations are possible because of the high Q of resonant lines at high frequencies and because of the low over-all interelectrode capacitance effect, since the individual interelectrode capacitances are connected in series. However, an adverse effect should be noted. The lead inductances are also effectively connected in series. This tends to cancel, somewhat, the beneficial effects of the series-connected interelectrode capacitances.

The power output, in general, of ring oscillators increases nearly in proportion to the number of pairs of tubes (assuming the same type of tubes is used). As in pushpull oscillators, the even harmonic content of the output is low.

34. Barkhausen-Kurtz Oscillator.¹

This u-h-f oscillator is of the positive-grid type. Wave lengths as low as a few centimeters have been obtained. The power output and efficiency are low, however.

The Barkhausen-Kurtz oscillator operates with the grid at high positive potential with the plate at a small negative potential. The connections are shown in Fig. 45. The action taking place may be explained in terms of the finite transit time of the electrons as they oscillate periodically past the grid in the region between the cathode and plate. The natural frequency of oscillation of the electron cloud is very high and is determined primarily by the potential of the grid. Electrons that lose energy to the field remain free to oscillate past the grid longer than electrons that absorb energy from the field. As a result, the average transfer of energy to the grid serves to maintain oscillations. The frequency observed at the electrodes of the tube depends on the natural frequency of oscillation of the electron cloud and upon the natural frequency of a coupled circuit (which may consist of the tube electrodes only). If the external circuit is in resonance or nearly so, it may greatly affect the observed frequency, as is usual with closely coupled circuits. The electronic oscillation may be considered as constituting the primary circuit. In this case, the oscillations are sometimes called Gill-Morell. According to the theory of coupled circuits, as applied by Wundt¹ to the Barkhausen-Kurtz oscillator, several coupling frequencies should be simultaneously possible, depending upon the damping of the coupled circuits. This has been verified by King,² who observed as many as three coupling frequencies maintained simultaneously.

It has been found that tubes in which Barkhausen-Kurtz oscillations may be produced usually have cylindrical electrodes, and Hollman³ has found that the ratio of plate-to-grid radii must be greater than 2 and less than 5, optimum values being between 2.5 and 3. The wavelength of the oscillating electron clouds is given approximately by the relation

$$\lambda^2 E_0 = K$$

¹ Wundt, R., *Hochfrequenztechn.* **36**, 133, 1930.

² King, R., *Rev. Sci. Instruments*, **11**, 615, 1940.

³ Hollman, H. E., "Physic und Technik der Ultrahurzen Wellen," Springer.

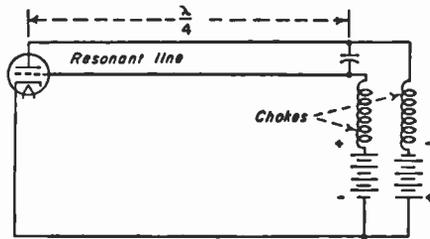


FIG. 45. Barkhausen-Kurtz oscillator employing resonant lines.

where λ = wavelength

E_g = grid potential

K = constant depending upon geometry of tube

35. Disk-seal Tube Oscillators.¹

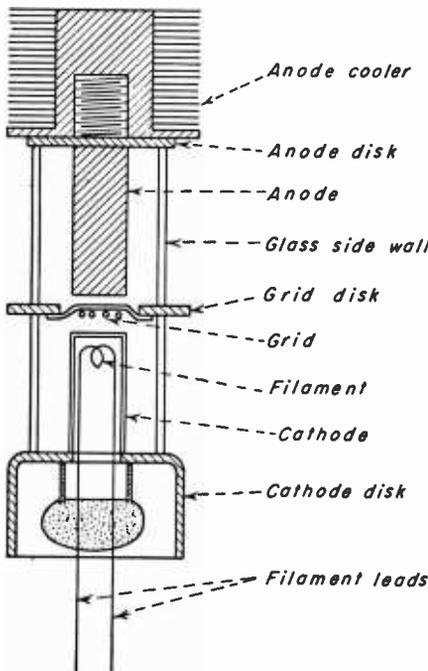


FIG. 46. Basic construction of disk-seal tube.

More recently, disk-seal type tubes² have been developed for operation in the u-h-f region. These tubes employ the space-charge control feature of conventional tubes but are so constructed that the elements, themselves, can form part of the bounding sides of cavity resonators. The basic mechanical construction of the disk-seal tubes may be observed in Fig. 46.

Whereas, in certain other oscillator circuits, the cathode functions as the common element between the plate and grid circuits, the disk-seal tube oscillator uses the so-called "grounded grid" circuit,³ wherein the grid becomes the common element. Advantages of this circuit lie in the fact that the interelectrode capacitances between grid and cathode, and grid and plate, are incorporated directly in the grid and plate circuits, respectively. The only coupling between the grid and plate circuits is in the plate-to-cathode capacitance, which is very small at the outset. Feedback thus may be easily controlled. Figure 47 shows a lumped-constant circuit of a grounded grid system.

The cavity resonators, of which the tube elements form a part, have the same effect at high frequencies as the lumped constants at lower frequencies. The interelectrode capacitances C_{pk} and C_{pk} are included in the distributed capacitance of the proper resonator. Since the electrical thickness of the cavity walls confines the electromagnetic fields to their respective cavities, no coupling exists between the two circuits, except for the small capacitance C_{pk} . Additional coupling must usually be added between grid and plate circuits before the system will oscillate. The cavity system equivalent of Fig. 47 is shown in Fig. 48. A typical disk-seal oscillator circuit is shown in Fig. 49.

Various methods of coupling and tuning of disk-seal cavity oscillator circuits are discussed by Gurewitsch,⁴ including the re-entrant type of oscillator, in which the grid cavity is mounted directly inside the plate cavity for coupling between the two circuits. Frequencies up to 3,300 Mc are typical of the reentrant

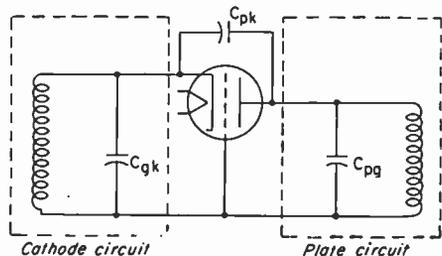


FIG. 47. Lumped-constant circuit for grounded grid oscillator.

¹ Gurewitsch, A. M., *Electronics*, February, 1946.

² McArthur, E. D., *Electronics*, February, 1945

³ Strong, C. E., *Electronics*, July, 1940.

⁴ *Op. cit.*

type of disk-seal tube oscillator. Owing to the relatively large amount of exposed metal surface, heat is readily dissipated and, as a consequence, considerable power may be handled.

36. Magnetron Oscillators.¹ The magnetron is essentially a diode, with coaxial electrodes, which is placed in a magnetic field, so that lines of electromagnetic force are approximately parallel to the axes of the diode electrodes. When the intensity of the magnetic field exceeds some critical value, the electrons will travel in orbits within the anode, and very few of them will reach it. When the intensity of the field is less than a certain value, all the electrons will reach the plate. Hence, the magnetic field can be used to control the anode current in a way similar to a grid in a triode. Originally, the magnetron oscillator was employed in this way, but its action was restricted to long wavelengths, and the tube could not compete successfully

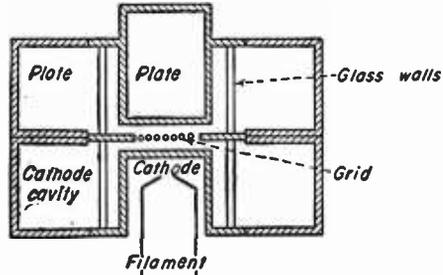


FIG. 48. Cavity system equivalent of Fig. 47.

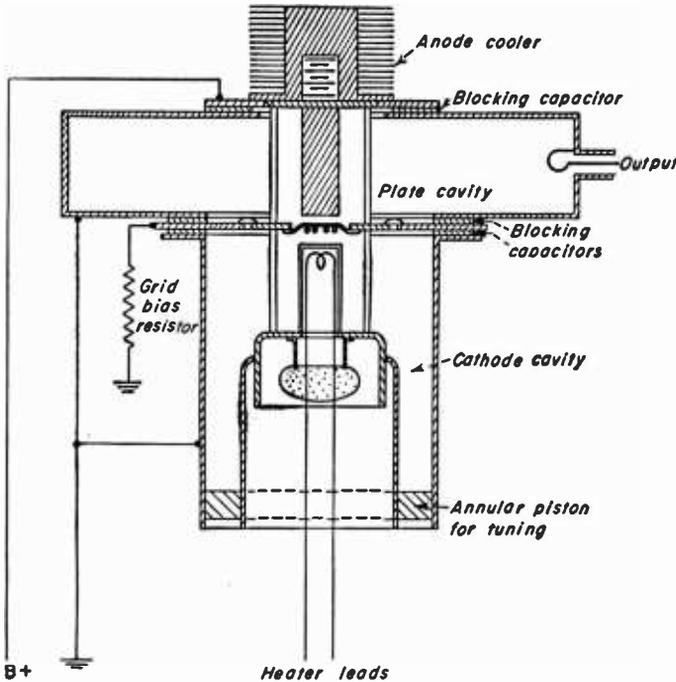


FIG. 49. Typical disk-seal-tube oscillator circuit.

with the triode oscillator. As a generator of u-h-f waves, however, the magnetron is superior to the triode. In its simplest form, the modern magnetron has its cylindrical

¹ Hull, A. W., *J. AIEE*, 40, 715, 1921; *Trans. AIEE* 42, 915, 1933; Elder, F. R., *Proc. IRE*, 13, 159, 1935; Megaw, E. C. S., *J. IEE (British)*, April, 1933; Kilgore, G. R., *Proc. IRE*, 24, 1140, 1936. Sarbacher, R. I., and W. A. Edson, "Hyper and Ultra-high Frequency Engineering." Wiley; Slater, J. C., *M.I.T. Radiation Lab. Rpts.* 43-9, August, 1942, and 43-28, March, 1943.

plate divided into two or more equal segments separated by narrow gaps, as shown in Fig. 50.

There are two distinct methods of producing h-f oscillations with magnetron tubes. These are (1) the negative-resistance (also called dynatron or quasi-stationary) method and (2) the electron-resonance (also called transit-time) method. The latter method may be further subdivided into the radial electron-coupled and rotational electron-coupled methods. With any of these methods, a constant magnetic field is used to control the direction rather than the magnitude of the electron current.

In the negative-resistance method, opposite ends of a center-tapped external-tuned circuit are connected to the split plate segments, as shown in Fig. 50. A negative resistance is developed, because the magnetic field causes electrons to be deflected along a spiral path and eventually fall on the segment which is at the lower instantaneous potential. The frequency of oscillation is substantially equal to the resonant frequency of the external circuit and, within its range of operation, independent of transit-time effects.

With the electron-resonance method, the transit time of the electrons is utilized to furnish oscillations. In the radial electron-coupled method, an external-tuned circuit

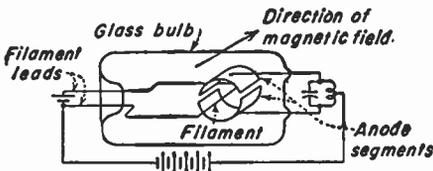


FIG. 50. Magnetron oscillator.

is used in conjunction with a solid, cylindrical plate. The external circuit is connected between the cathode and the plate. When the intensity of the magnetic field is adjusted to the critical cutoff value and the external circuit is tuned to the radial in-and-out frequency of the electrons about the lines of magnetic force, oscillations will be sustained by the transformation of part of the kinetic energy of the moving electrons into potential energy stored in the external circuit. The empirical relation $\lambda \cong 12,000/H$ (λ in centimeters and H in gauss) gives the approximate wavelength of the oscillations obtained.

In the rotational electron-coupled method, the external circuit, when used, is connected to split-plate segments. The electrons no longer return to the cathode, as in the above method, but flow to the plate along a path having a series of abrupt points. The motion of the electrons along these paths generates oscillations by inducing alternating charges on the segments as the electrons sweep by the segments. When the anode, itself, is a section of a resonant line (as distinguished from a resonant cavity), the segment may be coupled directly to a transmission line without the use of an external circuit. For microwave use, magnetrons are constructed with magnetically coupled resonant cavities as the anode structure. The cavities are excited to resonance by "bunches" of electrons forming a rotating, oscillating space charge around the cathode. The output is obtained by magnetic loop coupling, feeding into either a coaxial line or a wave guide.

The efficiency of a negative-resistance magnetron is from 16 to 65 per cent and outputs up to 1,000 watts continuous are available. Although very high frequencies may be obtained and the frequency range may be adjusted, the low efficiency of the radial electron-coupled magnetron precludes its use. A rotational electron-coupled magnetron has a reasonably high efficiency and relatively high power output, but has the disadvantage of limited frequency adjustment. Because of the construction of the anode in the resonant cavity version, large amounts of peak power, ranging in megawatts for some models, may be obtained. This factor is particularly useful for radar applications.¹

37. Klystron Oscillators.² Like the Barkhausen-Kurtz and electron-resonance magnetron oscillators, klystron oscillators employ the transit time of the electron

¹ An excellent discussion of cavity magnetrons may be found in L. Flanders, *Radio News*, Radio Electronic Engineering Edition, March, 1946, which includes a very good bibliography on magnetrons.

² References on klystron: Hansen, W. W., and R. D. Rechtmeyer, *J. Applied Phys.*, 10, 189, 1939;

stream to obtain u-h-f oscillations. Unlike tubes with conventional grid and/or plate structures, the klystron tube employs velocity modulation.

The action of velocity-modulated tubes may be explained as follows: An electron stream of constant density and speed is passed through a pair of grids. An oscillating electric field, parallel to the electron stream and of sufficient strength to alter the velocities of the electrons, is applied to the grids. After the electrons in the stream have passed through the grids, the electrons that have been accelerated begin to overtake the electrons that preceded them and were either (1) retarded, (2) not accel-

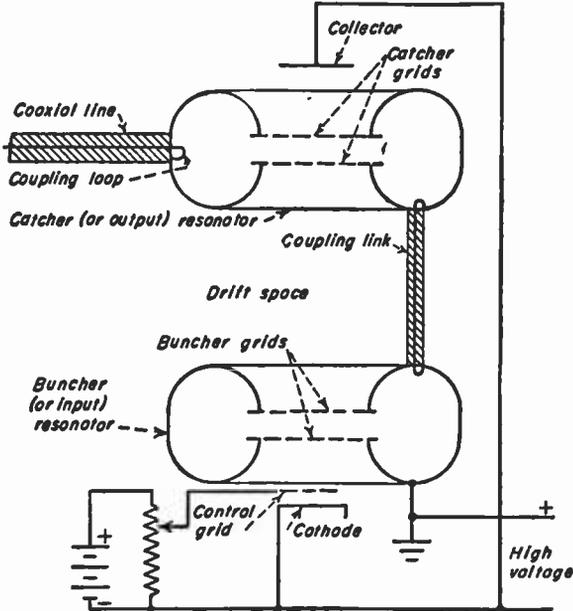


FIG. 51. Two-cavity klystron.

erated at all, or (3) not accelerated so much. This produces velocity modulation and groups the electrons into bunches, separated by relatively empty space. The electron bunches travel into another structure where they induce an output voltage and then pass on to a collector anode. By introducing part of the output power back into the input, oscillations may be sustained.

The modern two-cavity klystron, shown diagrammatically in Fig. 51, operates in essentially the same manner as described above. Resonant cavities, tuned to the desired frequency, provide the output- and input-tuned circuits with magnetic coupling between the two cavities. When the buncher resonator is excited, the buncher grids, as the name implies, perform the bunching operation on the electron stream. With the proper dimensions and operating voltages, the bunches of electrons pass on

Hansen, W. W., *J. Applied Phys.*, **9**, 654, 1938; Varian, R. H., and S. E. Varian, *J. Applied Phys.*, **10**, 321, 1939; Webster, D. L., *J. Applied Phys.*, **10**, 501, 1939; Heil, A. Arsenjewa, and O. Z. Heil, *Physik*, **95**, 752-762, 1935; Bruche, E., and A. Recknagel, *Z. Physik*, **106**, 459-482, 1938; Hahn, W. C., and G. F. Metcalf, *Proc. IRE*, **27**, 106-116, 1939; Hahn, W. C., *Gen. Elec. Rev.*, **42**, 258, 1939; Ramo, Simon, *Proc. IRE*, **27**, 757, 1939; Sarbacher, R. I., and W. A. Edison, *Proc. IRE*, **31**, 439-452, 1943; Muller, J. J., and E. Rostas, *Wireless Engr.*, **19**, 74, February, 1942; Sperry Gyroscope Co., Inc., *Klystron Technical Manual*, 1944 (this manual contains an excellent bibliography on klystrons); Harrison, A. E., "Klystron Tubes," McGraw-Hill, 1947.

through the drift space to the catcher, exciting the catcher resonator. Output may be obtained by coupling to the catcher resonator. Operation of the two-cavity klystron requires that both resonators be tuned to the same frequency. In actual

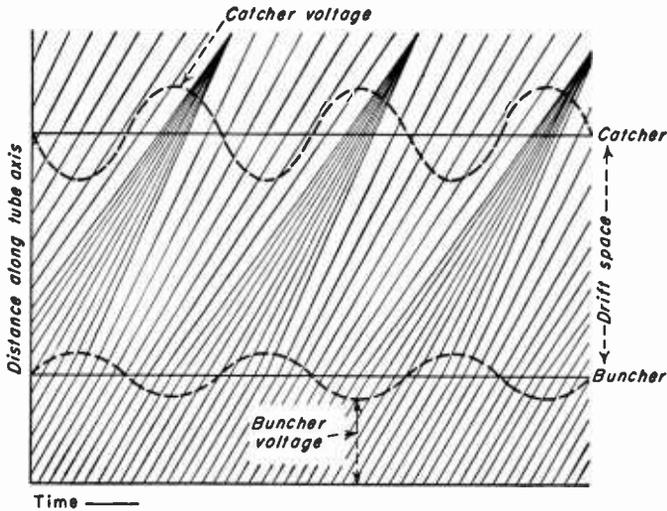


FIG. 52. Applegate diagram for two-cavity klystron.

models, tuning is accomplished either by changing the size of the cavities or by running tuning probes into the cavities.

The Applegate diagram, shown in Fig. 52, clarifies the action of the two-cavity klystron. Time is plotted as the abscissa and distance as the ordinate, the slope of lines representing the velocity of the electrons in the electron stream.

The multicavity klystron is constructed with several cavities along the drift space, each cavity being tuned to resonance. The effect of the additional cavities is to force stray electrons into either preceding or following bunches and thus increase the efficiency of the tube.

The reflex klystron has only one cavity but has a repeller anode which is maintained at a steady negative potential with respect to the cathode. The schematic diagram of the reflex klystron is shown in Fig. 53. The electrons with the greatest velocity come closest to the repeller anode before returning to the resonator grid structure. The electrons are alternately speeded up and slowed

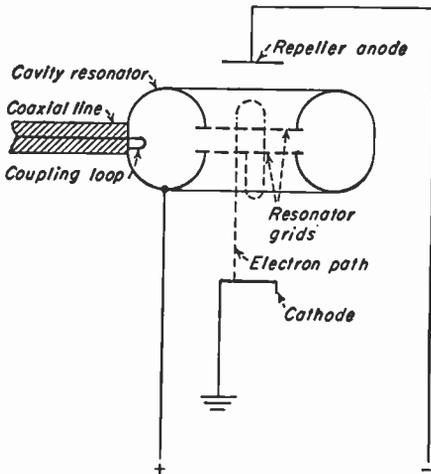


FIG. 53. Reflex klystron.

down in passing through the resonator grid structure, and bunching takes place in the space between the top resonator grid and the repeller anode. With appropriate voltages applied to the elements and with the cavity resonator properly adjusted, oscillations will be sustained. The frequency of oscillation may be varied by changing

the negative potential of the repeller anode and the dimensions of the cavity. The output and efficiency of reflex klystrons are small, being 80 mw and 0.7 per cent for a typical 10-cm reflex klystron.

38. Mechanical-electronic Oscillators. Mechanical-electronic oscillators are those which employ in combination both vacuum-tube circuits and mechanical-rotating members. The electrostatic audio generator developed by Kurtz and Larsen¹ falls into this class. It consists of a number of variable capacitors of the rotary type which are driven at a constant speed. When a direct voltage is connected through resistors to each of these capacitors, the charging current of the different capacitors is varied. If the plates of the capacitors are designed so that the charging currents are sinusoidal, a sinusoidal voltage will be developed across each of the series resistors. These voltages may be applied at the same time to the input of an amplifier. This generator was designed to produce a fundamental and 15 harmonic voltages simultaneously. The phase of any sinusoidal voltage with respect to the others is easily adjusted by shifting the position of any one of the stationary plates. To vary the amplitude of any of the

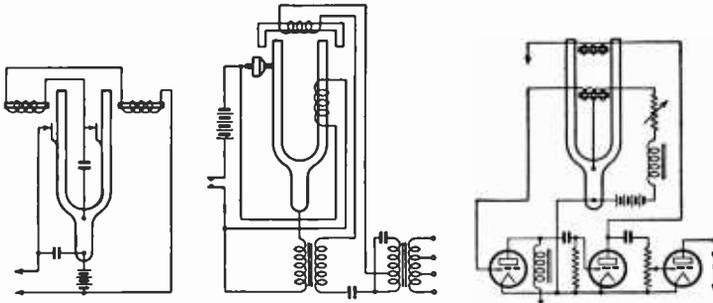


FIG. 54. Circuits for various types of tuning-fork oscillators (hummers).

voltages, it is necessary only to vary the particular applied direct voltage. A system such as this one is very useful in the study of the effects of changes in amplitude and phase of a complex sound on the ear.

The photoaudio generator developed by Schaffer and Lubszynsky² falls also into this class of oscillators. It consists of a system in which a beam of light falling on a phototube is interrupted by a perforated rotating disk. When a light aperture and the size and shape of the holes in the disk are correctly designed, the voltage developed by the photocell will be very nearly sinusoidal. The output of the cell can be suitably amplified and the frequency of the oscillator read by a tachometer applied to the driving motor.

39. Tuning-fork Oscillators. For oscillators of low to medium frequency the tuning fork provides an excellent resonator. The range from 100 to 10,000 cps is readily covered. Simplest and least precise are the contact-driven forks which are capable of supplying considerable power output of approximately square wave form. The single-button microphone drive gives a much purer wave and more constant frequency at the expense of power output. The double-button microphone drive gives a still purer wave, better stability, and greater output.

The free tuning fork, which is driven by one magnetic coil and excites another, is capable of high frequency stability. With a suitable circuit and a fork of special material, a stability of a few parts per million is obtained for relatively long times without benefit of voltage or temperature control.

40. Oscillator Automatic-frequency Control. It may be desirable in practice to control the frequency generated by an oscillator with a small change in voltage.

¹ Kurtz, E. B., and M. J. Larsen, *Elec. Eng.*, **54**, 950, 1935.

² Schaffer, W., and G. Lubszynsky, *Proc. IRE*, **19**, 1242, 1931; **20**, 363, 1932.

Special circuits which have been designed to do this are shown in Figs. 55a and b.¹ In Fig. 55b the tuned circuit of an oscillator is shunted by C and V_1 . Since the plate resistance of this tube depends upon the magnitude of the control voltage, the effective reactance of the combination is varied by variations in the control voltage. If the plate resistance is adjusted so that it is numerically equal to the reactance of C , the effective resistance does not change with a change in effective reactance. In this way,

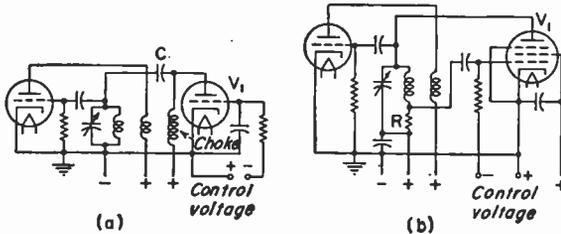


FIG. 55. Oscillators having automatic-frequency control.

a voltage applied to the control grid will control the frequency of oscillation without affecting the amplitude. Figure 55b shows another type of control circuit for varying the frequency of an oscillator. Here, the plate circuit of the control tube V_1 shunts the resonant circuit of the oscillator. The voltage drop in R is applied to the grid of the control tube. This voltage is approximately 90 deg out of phase with the resonant circuit voltage, and hence the plate circuit of the control tube affects the tank circuit in the same way as a reactance. Variations in this effective reactance are obtained by changing the bias on the control tube.

41. Electron-coupled Oscillators. It has been pointed out that the frequency of oscillation will be affected by changes in load unless the current in the resonant circuit inductance is not changed when the load is varied. The use of buffer amplifiers between the oscillator and the load aids in shielding the former from load variations.

Dow² has developed another method of making the frequency of oscillation independent of load variations. The method employs electron coupling between the oscillator and the load. This consists of a Hartley oscillator in which the screen grid serves as anode while the plate serves only as an output electrode (see Fig. 56). The screen grid is effectively grounded to alternating currents while the cathode is at an alternating potential above ground. This prevents the load impedance in the plate circuit from reacting back on the oscillator. At the same time the electrons that pass through the screen are attracted by the more positive plate. This results in the plate current having an alternating component which is of the same frequency as the oscillator frequency. Hence energy is delivered to the load through the electron stream, and at the same time the oscillator is effectively shielded from the load.

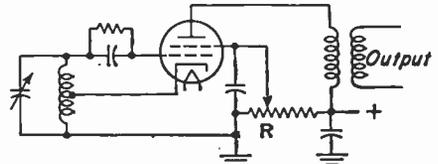


FIG. 56. Electron-coupled oscillator.

The frequency of oscillation with this type of oscillator can be made independent of supply voltage variations by properly choosing the ratio of screen grid to plate potential. This is possible because there is always some value of this ratio for which the frequency is independent of the applied voltage. By adjusting the position of the tap on R until the frequency becomes independent of applied voltage, a high degree of frequency stability may be obtained. It may be necessary to change the position of this tap if the oscillator frequency is varied over wide limits.

When the oscillator is operated normally, the tube is biased beyond cutoff, and the plate current flows in pulses that are very rich in harmonic content. Since the frequency stability is very high and the output extremely rich in harmonics, this oscillator

¹ Travis, C., *Proc. IRE*, **23**, 1125, 1935. For variants of the basic circuit, see Freeman, R. L., *Electronics*, p. 20, November, 1936; Foster, D. E., and S. W. Seeley, *Proc. IRE*, **25**, 289, 1937.

² Dow, J. B., *Proc. IRE*, **19**, 2095, 1931.

is very satisfactory for use in heterodyne frequency meters. There are many possible ways in which the principle of electron coupling may be used to great advantage.¹

42. Power-oscillator Design. In applications which require appreciable power output and for which a high degree of frequency stability is not necessary, feedback oscillators may be used. In these oscillators the tube operates as in a class C amplifier, and oscillators adjusted in this way are referred to as class C oscillators.

The design of a class C oscillator may be conducted along the same lines as for a class C amplifier. There are two principal points of difference: (1) the power required to supply the energy for the grid circuit must be obtained from the plate-circuit power supply since this, with the exception of the filament supply, is the only source of power in the circuit; and (2) the oscillating circuit must contain sufficient stored energy to meet the grid-circuit requirements. The first step in design is to determine the correct operating conditions for the tube. These may be determined by various methods, some of which consist of graphical integrations of the current waves that are obtained from the static characteristic curves of the tube.² These methods require that the complete static characteristic curves be available far out in the region of positive grid potential. To obtain these curves special experimental techniques have been devised.³ Other methods consist of analytical integration of (1) simple expressions which are assumed to approximate the wave form of the current pulses or (2) approximate analytical expressions for the static characteristic curves.⁴ Among these methods of precalculation, some are extremely laborious but accurate, while others are rapid but inaccurate. The methods due to Chaffee⁵ and to Terman and Roake⁶ are recommended as sufficiently accurate for engineering purposes, and at the same time they are reasonably rapid. In addition to these methods of precalculation, the operating conditions for the tube may be determined by direct test.

The results obtained by any one of these various methods will be subject to certain limitations which are always imposed on the operation of the tube. These limitations are designed to preserve reasonable life and to prevent sudden failure of the tube. Electrode dissipations are limited to prevent liberation of gas or injury to the tube structure due to melting or warping. The maximum polarizing potential is limited to prevent flashover due to gas or cold emission, breakdown of insulation, or puncture of the tube due to stray beams of electrons. Also, the grid dissipation must be kept small enough to avoid appreciable primary emission from the grid, and the maximum instantaneous current flowing in the tube must be limited to preserve reasonable filament life. When the operating conditions are selected with due regard to these limitations and to the conditions of power output and efficiency desired, they may be used for the design of any one of the oscillator circuits shown in Fig. 1.

These operating conditions are as follows:

E_B = direct plate voltage

E_c = direct grid voltage

E_{p1} = rms value of fundamental component of plate alternating voltage

E_{g1} = rms value of fundamental component of grid alternating voltage

I_B = average value of direct plate current

I_c = average value of direct grid current

¹ EDITOR'S NOTE. Very good stability can be obtained from a Colpitts oscillator modified by placing the tuning capacitor in series with the inductance and by using the maximum possible fixed grid-cathode and cathode-ground capacitances and the maximum possible inductance of high Q . A pentode may be employed as an electron-coupled oscillator in this Colpitts arrangement.

² Prince, D. C., *Proc. IRE*, 11, 275, 405, 527, 1923; Kilgour, C. E., *Proc. IRE*, 19, 42, 1931; Mourontseff, I. E., *Proc. IRE*, 20, 783, 1932; Mourontseff, I. E., and Kozanowski, H. N., *Proc. IRE*, 22, 1090, 1934; *Proc. IRE*, 23, 752, 1935; Wagener, W. G., *Proc. IRE*, 25, 47, 1937.

³ Kozanowski, H. N., and I. E. Mourontseff, *Proc. IRE*, 21, 1082, 1933; Chaffee, E. L., *Electronics*, June, 1938.

⁴ Fay, C. E., *Proc. IRE*, 20, 548, 1932; Everitt, W. L., *Proc. IRE*, 22, 152, 1934; Terman, F. E., and J. H. Ferns, *Proc. IRE*, 22, 359, 1934; Miller, B. F., *Proc. IRE*, 23, 496, 1935; Babits, V. A., *L'Onde électrique*, 14, 668, 1935; Everitt, W. L., *Proc. IRE*, 24, 305, 1936; Terman, F. E., and W. C. Roake, *Proc. IRE*, 24, 620, 1936; Wagener, W. G., *Proc. IRE*, 25, 47, 1937.

⁵ Chaffee, E. L., *J. Applied Phys.* 9, 471, 1938.

⁶ Terman, F. E., and W. C. Roake, *Proc. IRE*, 24, 620, 1936.

I_{p1} = rms value of fundamental component of plate a.c.
 I_{g1} = rms value of fundamental component of grid a.c.

43. Power Relations in Class C Oscillators. The d-c power supplied to the oscillator circuit from the source E_B is

$$P_{\text{input}} = I_B E_B$$

The power output to the tank circuit at fundamental frequency is

$$P_{\text{tank}} = E_p I_p$$

This power must supply the driving power, the load, and the losses in the tank circuit. This power may also be expressed as

$$P_{\text{tank}} = I_L^2 R_L$$

where R_L represents, in addition to the inherent resistance of the total tank inductance, the resistance reflected into the tank circuit by the load requirement of the grid circuit and by the load itself. I_L is the circulating current in the oscillating circuit.

The power lost at the plate is

$$P_{\text{plate}} = I_B E_B - E_p I_p$$

The driving power required by the tube is

$$P_{\text{driving}} = E_g I_g$$

This power supplies the power delivered to the grid resistor to maintain the bias voltage

$$P_{\text{grid resistor}} = I_c E_c$$

and the power lost at the grid. Therefore, the grid loss is

$$P_{\text{grid}} = E_g I_g - I_c E_c$$

The power available for output is then given approximately by

$$P_{\text{output}} = E_p I_p - E_g I_g$$

Actually, the useful power which may be obtained from the oscillator will be less than is given by this expression, by the inherent losses in the circuit elements. However, in a well-designed oscillator these may be kept small, and the power output will be very nearly that given above; hence the efficiency of the oscillator may be expressed as

$$\text{Eff} = \frac{E_p I_p - E_g I_g}{I_B E_B}$$

This is not the over-all efficiency. An expression for the over-all efficiency of an oscillator may include the power required for the filament and the losses incurred in obtaining the high-voltage plate supply.

44. Design of a Hartley Oscillator. The methods illustrated here are applicable to any of the other feedback oscillators. The problem is to determine the values of the circuit elements in such a way that the proper voltages (which have been obtained by any one of the various methods discussed above) will be applied to the tube. The alternating plate and grid voltages should be 180 deg out of phase for correct operation. Most of the methods for obtaining the operating conditions assume that this phase condition will be met. Slight variations in phase of a few degrees have only a small effect on the performance of the oscillator.

In the Hartley circuit the voltages developed across the tapped inductances are used as the alternating components of the plate and grid voltage. The total alternating voltage across the tank circuit is the sum of E_p and E_g . The filament tap is adjusted so that the ratio of the plate and grid voltage is E_p/E_g . It has been found experimentally that the effective selectivity Q of the tank should be greater than 12.5, approximately. Higher values of Q increase the stability and lower the harmonic

content. The tank-circuit inductance for any frequency may be found by use of the relation

$$L = \frac{(E_{p1} + E_{g1})^2}{2\pi f Q P_{\text{tank}}}$$

and the tank-circuit capacitance by

$$C = \frac{QP}{2\pi f (E_p + E_g)^2} \quad C = \frac{QP_{\text{tank}}}{2\pi f (E_{p1} + E_{g1})^2}$$

where f is the required frequency of oscillation.

The grid-bias voltage required is given by E_c , and with an average grid current of I_c , the grid resistance is

$$R_c = E_c / I_c$$

The value of the grid capacitor C_c should be large enough to act as an effective short circuit for the grid resistance R_c at the frequency of operation. It should not, however, be so large as to cause intermittent oscillations, as discussed in Sec. 14 of this chapter.

The plate blocking capacitor C_B should be large enough so that the reactive voltage developed across it will be small in comparison with E_p . The inductance of the shunt feed choke should be large in comparison with the inductance of the portion of the tank circuit which it effectively shunts. In the Hartley oscillator this blocking capacitor and choke can be designed in such a way as to help correct for variations in phase of the alternating plate and grid voltages from the 180-deg position.

45. Series-tuned Colpitts Oscillator. If, in the Colpitts oscillator of Fig. 1, C_p and C_p are fixed in value and are made very large compared with the tube capacitances, and if the tuning capacitance is inserted in series with the inductance, the stability of the oscillator compared with a conventional Colpitts circuit will be materially improved. For maximum stability the inductance should be large and of high Q . A discussion of this circuit will be found in the Clapp and Sandeman references cited below.

References

Oscillators, General:

- Arguimbau, L. B.: *Proc. IRE*, **21**, January, 1933.
- Benioff, H.: *Proc. IRE*, **19**, July, 1931.
- Brunetti, C.: *Proc. IRE*, **25**, 1595, 1937.
- Clapp, J. K.: *Proc. IRE*, **26**, 356, March, 1948.
- Dow, J. B.: *Proc. IRE*, **18**, May, 1927; **19**, December, 1931.
- Eller, K. B.: *Proc. IRE*, **16**, December, 1928.
- Groszkowski, Janusz: *Proc. IRE*, **21**, 958, 1933; **22**, February, 1934.
- Hamburger, F., Jr.: *Proc. IRE*, **22**, January, 1934.
- Herold, E. W.: *Proc. IRE*, **23**, 1201, 1935. (Contains large bibliography on negative resistance oscillators.)
- Horton, J. W.: *Bell System Tech. J.*, **3**, 508, July, 1924.
- Jimbo, S.: *Proc. IRE*, **17**, November, 1929.
- King, R. W.: *Bell System Tech. J.*, **2**, October, 1923.
- Le Corbeiller, P.: *J. IEE, Wireless Sec.*, **11**, 292, 1936.
- Model, S. J.: *Proc. IRE*, **21**, December, 1933.
- Rosenstein, H. O.: *Proc. IRE*, **19**, October, 1931.
- Sandeman, E. K.: "Radio Engineering," p. 421, Wiley, 1948.
- Shenk, E. R.: *Electronics*, January, February, March, 1944.
- Stone, J. S.: *Proc. IRE*, **19**, August, 1931.
- van der Pol, Balth: *Proc. IRE*, **22**, September, 1934.
- Vecchiacchi, F.: *Proc. IRE*, **19**, May, 1931.

Frequency-control Crystals:

- Boella, M.: *Proc. IRE*, **19**, July, 1931.
- Cady, W. G.: *Proc. IRE*, **10**, April, 1922.
- Clayton, J. M.: *QST*, September, 1926.
- Colwell, R. C.: *Proc. IRE*, **20**, May, 1932.
- Dye, D. W.: *Proc. Phys. Soc. (London)*, **28**, August, 1926.

- Harrison, J. R.: *Proc. IRE*, December, 1927; 16, November, 1928.
 Heaton, V. E., and E. G. Lapham: *Proc. IRE*, 20, February, 1932.
 High, S. C., and G. W. Willard: *Proc. IRE*, 25, May, 1937.
 Horton, J. W., and W. A. Marrison: *Proc. IRE*, 16, February, 1928.
 Koga, I.: *Proc. IRE*, 19, June, 1931.
 Lapham, E. G.: *Proc. IRE*, 22, January, 1934.
 MacKinnon, K. A.: *Proc. IRE*, 20, November, 1932.
 Mason, W. P.: *Bell System Tech. J.*, 19, January, 1940.
 Meahl, H. R.: *Proc. IRE*, 22, June, 1934.
 Mueller, E. L.: *QST*, May, 1927.
 Pierce, G. W.: *Proc. Am. Acad. Arts Sci.*, 59, October, 1923.
 Terry, E. M.: *Proc. IRE*, 16, November, 1928.
 Van Dyke, K. S.: *Proc. IRE*, 16, June, 1928.
 Vigoureux, P.: *Phil. Mag.*, December, 1928.
 Watts, W. A.: *QST*, January, 1928.
 Wheeler, L. P.: *Proc. IRE*, 19, April, 1931.
 Williams, N. H.: *Proc. IRE*, 21, July, 1933.
 Worrall, R. H., and R. B. Owens: *Proc. IRE*, 16, June, 1928.
 Wright, J. W.: *Proc. IRE*, 17, January, 1929.

High-frequency Oscillators (including magnetrons, Klystrons, electron coupling, etc.):

- Barkhausen and Kurz: *Physik. Z.*, 21, 1920.
 Brillouin, L.: *Elec. Commun.*, 20, 1941.
 Dytrt, L. F.: *Proc. IRE*, 23, March, 1935.
 Gill and Morell: *Phil. Mag.*, 44, 1922.
 Hoag, J. B.: *Proc. IRE*, 21, August, 1933.
 Hollmann, H. E.: *Proc. IRE*, 17, February, 1929.
 Karplus: *Gen. Radio Experimenter*, 5, May, 1931.
 Kilgore, G. R.: *Proc. IRE*, 20, November, 1932.
 Kompfner, R.: *Wireless Engr.*, 17, November, 1940.
 Kozanowski, H. N.: *Proc. IRE*, 20, June, 1932.
 Loeber, C. W.: *Electronics*, November, 1930.
 McNamara, F. T.: *Proc. IRE*, 22, August, 1934.
 Megaw, E. C. S.: *Proc. IRE*, 21, December, 1933.
 Moore, W. H.: *Proc. IRE*, 22, August, 1934.
 Mouromtseff, I. E., and H. V. Noble: *Proc. IRE*, 20, August, 1932.
 Okabe, Kinjiro: *Proc. IRE*, 21, November, 1933.
 Thompson, B. J., and P. D. Zottu: *Proc. IRE*, 22, December, 1934.
 Webster, D. L.: *J. Appl. Phys.*, 10, December, 1939.
 Wenstrom, W. H.: *Proc. IRE*, 20, January, 1932.
 White, W. C.: *Electronics*, April, 1930.

CHAPTER 11

MODULATION AND DETECTION

BY LESLIE F. CURTIS¹

TYPES OF MODULATION

1. Modulated Waves. A *modulated wave* is a periodic wave of which the amplitude, frequency, or phase is varied in accordance with a signal. *Modulation* is the process by which this variation is accomplished. *Demodulation* or *detection* is the process by which the original signal is recovered from the modulated wave.

The unmodulated component of the original wave is a *carrier wave* or, more broadly, a *carrier*. The frequency of the carrier usually is much higher than the highest frequency component in the original signal.

The *modulating wave* or *signal wave* applied to the carrier may be a direct representation of the original signal, or it may be a wave of a different carrier frequency previously modulated by the original signal. In this case the process is called *double modulation*, and the modulation appears in new frequency groups associated with new carrier frequencies equal to the sum and difference of the subcarrier and fixed-carrier frequencies. Double modulation accompanied by the selection of one of the new carrier frequencies to which the signal modulation has been transferred is often called *frequency conversion*.

In tubes and circuits the waves are identified by the current in, or the voltage across, circuit elements. Either current or voltage may be used as a reference since their values are related by the energy loss or energy storage in the circuit elements. In linear elements the proportionality is expressed by the circuit impedance or admittance.

A modulated wave of current is expressed by

$$i = I \cos(\omega t + \phi) \quad (1)$$

where I = amplitude

$\omega/2\pi$ = frequency of the carrier

t = time

ϕ = relative phase

The choice of a sine or cosine function for representing the conditions is a matter of convenience. The signal may be imparted to the carrier either by a variation of the amplitude or by a variation of the phase as a function of time. Variations in phase are accompanied by simultaneous variations in frequency, since the frequency is $1/2\pi$ times the derivative of the phase with respect to time. Simultaneous variations of phase and amplitude during modulation usually lead to distortion.

Amplitude modulation (a.m.) is the process by which the amplitude of the wave is controlled by a signal for the transmission of intelligence. Only the value of I in Eq. (1) is varied. In *pulse modulation* the amplitude is increased only during definite intervals, repeated in a definite time sequence, for only a small fraction of the total time. It is a distinctive type of amplitude modulation and by general acceptance is classed separately.

Velocity modulation is a process in which the velocity of electrons in transit in a special tube is controlled by a signal. The wave at the output terminals of the tube is modulated in amplitude. This process is usable only at extremely high carrier frequencies.

¹ Hazeltine Electronics Corporation.

Frequency modulation (f.m.) and *phase modulation (p.m.)* are processes by which the instantaneous frequency and, therefore, the phase of the wave relative to that of the carrier are controlled by the signal for the transmission of intelligence. Only the value of ϕ in Eq. (1) is varied. The *deviation* either of the frequency or of the phase from the reference frequency or phase of the carrier may be made proportional to the signal, thereby identifying the modulation with one of the above types.

2. Amplitude Modulation. A current wave amplitude modulated by a single signal component of frequency $a/2\pi$ may be expressed by rewriting Eq. (1) as

$$i = I_0(1 + m \cos at) \cos(\omega t + \theta) \quad (2)$$

where I_0 = amplitude of carrier

m = relative variation in amplitude at signal frequency

θ = constant phase of carrier

The value of m is called the *degree of modulation* or *modulation factor*. When multiplied by 100, it is the *percentage of modulation*.

The phase of the carrier may be neglected unless the current is to be combined with other currents of the same frequency. Neglecting θ and making I_0 unity for convenience, Eq. (2) may be expanded to

$$i = \cos \omega t + \frac{m}{2} \cos(\omega + a)t + \frac{m}{2} \cos(\omega - a)t \quad (3)$$

The modulated wave contains, in addition to the carrier, two independent periodic waves spaced therefrom in frequency by the modulating frequency. These are *modulation side frequencies*.

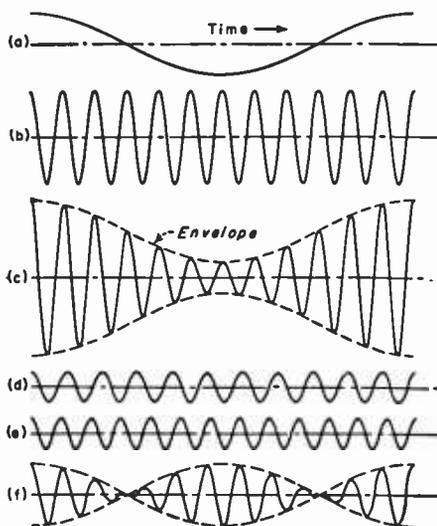


FIG. 1. Components in a-m waves.

These waves are illustrated in Fig. 1, in which a is the signal, b the unmodulated carrier, c the complete modulated wave, d the lower side frequency, e the upper side frequency, and f the two side frequencies without the carrier. The *envelope* of the composite wave (shown dotted) has the same shape as the original signal wave. The intercepts of the composite wave with the zero axis are not changed by a.m.

3. Amplitude Modulation with Several Signal Frequencies. The modulation may be expressed as a Fourier series when the original signal wave is continuous though it contains components at several frequencies. The composite signal is then

$$i = I_0[1 + \sum m_n \cos(nat + \alpha_n)] \cos \omega t \quad (4)$$

where $\alpha/2\pi$ = highest common factor of signal frequencies

m_n = variation in amplitude at the n th harmonic signal frequency relative to carrier

α_n = phase of n th signal harmonic

When Eq. (4) is expanded, independent pairs of side frequencies appear for each signal frequency ($n\alpha/2\pi$). The bands of frequencies above and below the carrier frequency occupied by the side frequencies are *upper side bands* and *lower side bands*, respectively. The *communication band width* is the frequency spectrum occupied

by both side bands and carrier. In a.m. it is two times the highest modulating frequency in the original signal.

The relative phase (α_n) of each of the signal components must be preserved to maintain the original form of the signal. This is relatively unimportant in signals for music and speech but is sometimes exceedingly critical, as in video signals in television systems.

Inward (or *downward*) modulation and *outward* (or *upward*) modulation are the respective decreases or increases in the envelope of the composite signal wave relative to the carrier. Inward and outward modulation are not necessarily alike. Inward modulation must not reduce the carrier to zero, or the character of the original signal will be lost. The maximum outward modulation determines the maximum power in the modulated wave.

In many cases it is sufficient to express the complete wave as

$$i = I_0(1 + M) \cos \omega t \quad (5)$$

when the instantaneous modulation M varies slowly with respect to the frequency of the carrier.

Carrier suppression is the process of balancing out the carrier component in an a-m wave, leaving only the frequency components in the side bands (see Fig. 1f). The transmitted power is then zero in the absence of modulation, resulting in an increase in transmitting efficiency. A carrier must then be supplied locally at the receiver for detection.

Single-side-band and *vestigial-side-band* signal waves are a-m waves in which all or a portion of the side-frequency components above or below the carrier frequency has been removed by suitable filters. The portion of the frequency spectrum from which the components were removed is then available for other services.

While pulse modulation and the complete video modulation due to the camera and synchronizing signals in television are, within the broad definition, types of a.m. by common usage an "a-m station" is understood to be a transmitting station in which the amplitude of the carrier is linearly and uninterruptedly controlled in accordance with a signal, as, for example, music or speech. The average value of the transmitted wave is then equal to that of the carrier, and the average inward and the outward modulation are equal.

The signal frequencies for a-m broadcasting are ordinarily in the range from 30 to 8,000 or 10,000 cycles; both side bands are transmitted, and the communication band width is from 15,000 to 20,000 cycles. The maxima are about half as great in commercial radiotelephony. The signal frequencies are confined to the range from about 150 to 3,000 cycles in some special services where the intelligibility of speech in the presence of interference is of more consequence than fidelity.

In multiple-channel radio transmission, separate subcarriers amplitude-modulated by the several telephone or telegraph signals may be applied to provide single-side-band modulation for a final carrier. The communication band width is then equal to the product of the number of channels and the frequency spacing of the subcarriers.

The wave that transmits the video signal in television may be considered to be subject to outward pulse modulation for the synchronizing signals and to inward amplitude modulation for the camera signals, both referred to the blanking level. (See chapter on Television for video wave form.) The pulses are in a definite sequence which is repeated at the frame frequency and at uniform amplitude. The inward modulation varies with the brightness of the individual picture elements and may reduce the instantaneous wave to any level that may be transmitted satisfactorily without interference. It may include significant signal frequencies as high as 4 Mc in scanning consecutive picture elements. The average of the transmitted wave differs for bright or dark pictures and may be constant, as with still pictures, or may vary with changes in scene. The system should, therefore, be capable of handling signal frequencies from zero to about 4 Mc. Vestigial-side-band

transmission is used for the video signal to conserve space in the frequency spectrum for other assignments. A complete television channel also includes the associated f-m sound carrier.

4. Pulse Modulation. A pulse-modulated wave is reduced to zero in the intervals between pulses, which occupy only a small portion of the total time. The envelopes of the transmitted pulses are somewhat trapezoidal, although desirably as nearly rectangular as possible. While intelligence may be transmitted by controlling the amplitude, width, or spacing of the pulses in accordance with a signal, only width or spacing control is normally used, since it is then possible to utilize maximum power during each pulse and to depend on the time position of the leading and trailing edges of the pulses for the required intelligence. These edges are relatively independent of amplifier nonlinearities, and in many cases the rectified pulse may be clipped at the receiver to eliminate the amplitude effects of noise, so that the accuracy of reproduction and the signal-to-noise ratio are better than for conventional a-m systems. When the amplitude of the transmitted wave is equal to or greater than the noise, the improvement over a.m. is roughly proportional to the

increase in the transmission band width required to reproduce the pulse-type signal.

It is convenient to dimension pulses in microseconds although time in equations appears in seconds. Modulation with a single pulse per repetition cycle is illustrated in Fig. 2, in which the signal is shown at (a) and the modulated wave at (b). The leading edges of the pulses are separated by the period t_r , which is $1/f_r$, where f_r is the pulse-repetition rate or pulse-repetition frequency (p.r.f.). The ratio of the average power during the repetition cycle to the average power during the

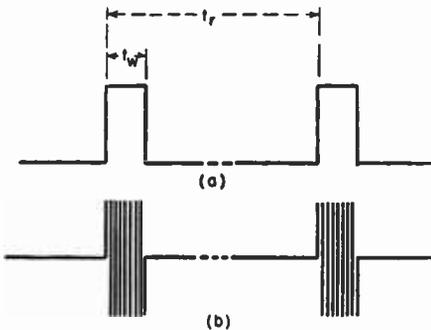


Fig. 2. Single-pulse modulation.

pulse (sometimes called the "peak power") is called the *duty* or *duty cycle* and gives a measure of the average heating in the circuit in comparison with the pulse requirements which determine both maximum voltage and current. The actual peaks of power during the r-f cycles are twice the pulse power. The duty cycle with the usual approximately rectangular pulse is

$$\frac{\text{Average power}}{\text{Pulse power}} = \frac{t_w}{t_r} = f_r t_w \quad (6)$$

where t_w = pulse width

Single-pulse transmission has been used extensively in radar, in interrogator units for radio beacons (racons) for ships and aircraft, and in the transmitted reply of these beacons. The transmission may be interrupted according to an audible code at infrequent intervals (slow coding), or different pulse widths may be used at times (fast coding), to transmit a limited amount of intelligence. The pulse width or the pulse spacing may be controlled in accordance with the amplitude of an audio signal, thereby providing *pulse-width modulation* or *pulse-time modulation* for a communication channel, although such arrangements with single pulses do not have the accuracy, flexibility, or channel capacity of multiple-pulse systems. Pulse-time modulation is also called *pulse-position modulation*.

Pulse widths from 0.1 to 100 μ sec have been used in radar and racon transmission with repetition rates from 60 to 5,000 per sec, depending on the requirements. Pulse power in a radar transmitter is usually in excess of 20 kw while in racon transmitters it may be only a few watts.

Since actual r-f pulses are only approximately rectangular, their shapes are

usually described in terms of width, rise time, and decay time of their envelopes. An exaggerated pulse shape is shown in Fig. 3. *Rise time* is the time in microseconds for the pulse to increase from some specified low value to a nearly uniform value; e.g., from 10 to 90 per cent of peak amplitude. When, because of circuit or tube characteristics, the rise ends in a very sharp and short peak or *spike*, the latter is disregarded in measuring amplitude. *Decay time* is rated by a similar measurement. Pulse width is measured at some specified fraction of full amplitude; e.g., 50 per cent. The highest signal frequency contained in a pulse is roughly one-half the reciprocal of the rise or decay time, whichever is shorter. Both side bands are transmitted, but demodulated pulse shapes at the output of a receiver are often degraded by insufficient circuit band width.

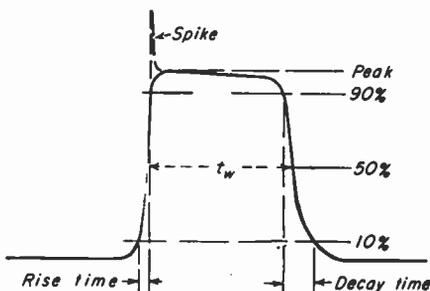


FIG. 3. Exaggerated pulse.

A modulation envelope with multiple pulses is illustrated in Fig. 4, in which the signal includes pulses following the first or marking pulse in a group. Additional intelligence may be transmitted by the extra pulses depending on their spacing, number, and width. Some of the possibilities that exist in such combinations are

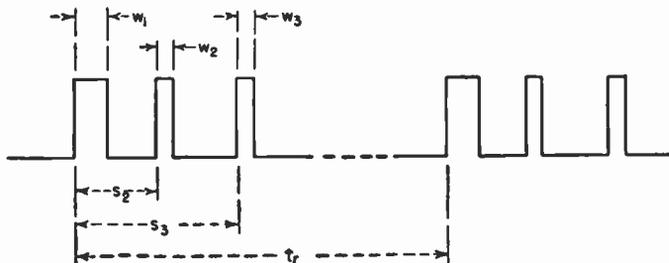


FIG. 4. Multiple-pulse modulation signal.

illustrated by the example that follows. An eight-channel, time-division, multiplex, *pulse-position* modulated, u-h-f communication system^{1,2} utilizes 2- μ sec synchronizing pulses at a repetition rate of 10,000 per sec to provide a frame of 100 μ sec. The time remaining in each frame is divided into eight intervals approximately 12 μ sec

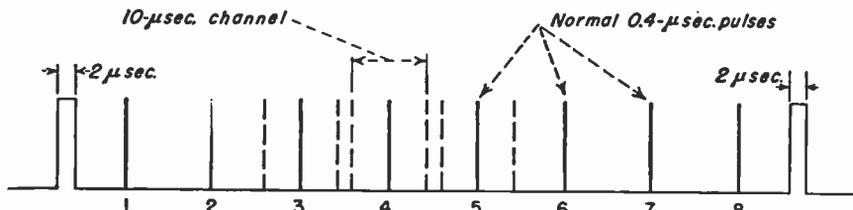


FIG. 5. Pulses in eight-channel pulse-position modulated time-division multiplex signal.

long, one being assigned to each signal channel as illustrated in Fig. 5. The channels are switched in succession at appropriate times relative to the synchronizing pulse

¹ Kelleher, J. J., Pulse-modulated Radio Relay Equipment, *Electronics*, 19, 124-129, May, 1946.
² Greer, W. R., Pulse Modulating System, *Electronics*, 19, 126-131, September, 1946.

at both transmitter and receiver. The transmitted wave for each signal channel is modulated by a 0.4- μ sec pulse, which is controlled in position relative to the center of the assigned channel interval by the amplitude of the signal for that channel. The displacement is about 5 μ sec either side of normal for maximum modulation, leaving a 2- μ sec guard interval between channels. Each signal is sampled 10,000 times per second, and the maximum signal frequency that may be transmitted with good fidelity is about one-third of this frequency, or 3,000 cycles. Voice, telegraph, or facsimile signals are handled satisfactorily.

5. Phase Modulation and Frequency Modulation.

The peak amplitude of the composite signal is constant in p-m and f-m waves. The signal is imparted to the carrier by a variation of the phase as a function of time.

A single signal component of frequency $a/2\pi$ produces a current

$$i = I_0 \cos(\omega_c t + \theta + m \sin at) \quad (7)$$

where I_0 = constant amplitude

$\omega_c/2\pi$ = constant carrier frequency

θ = constant relative carrier phase

m = relative maximum variation in phase or modulation index due to the signal

θ may be neglected unless the current is to be combined with other currents of the same frequency.

The instantaneous angular frequency ω of the composite wave is the derivative of the instantaneous phase with respect to time and is

$$\omega = \omega_c + ma \cos at \quad (8)$$

The peak frequency deviation Δf is $a/2\pi$ times the peak phase departure or modulation index m in radians and is

$$\Delta f = \frac{ma}{2\pi} = \frac{\Delta\omega}{2\pi} \quad (9)$$

The original signal usually contains components at several frequencies. The maximum phase departure and the

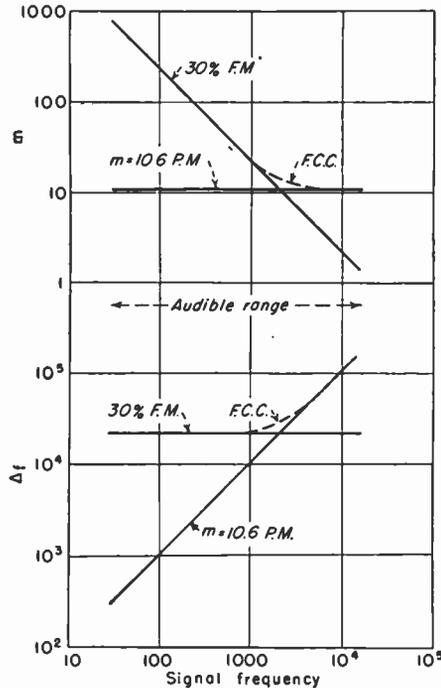


FIG. 6. Variation of m and Δf with signal frequency.

maximum frequency deviation for several unrelated modulating frequencies are then

$$m = m_1 + m_2 + m_3 + \dots + m_k \quad (10)$$

and

$$\Delta f = \frac{1}{2\pi} (m_1 a_1 + m_2 a_2 + \dots + m_k a_k) \quad (11)$$

When the component maximum phase departures m_1 to m_k are made proportional to the amplitudes of the signal components at frequencies $a_1/2\pi$ to $a_k/2\pi$, the composite wave is said to be *phase-modulated*.

When the component maximum frequency deviations are made proportional to the amplitudes of the signal components, the composite wave is said to be *frequency-modulated*.

The variations in m and Δf with signal frequency are shown in Fig. 6 for both p.m. and f.m. as indicated. The dashed curves show the variations in these quantities

Illustrative curves for $J_n(m)$ for constant n and varying m are shown in Fig. 8. The higher order functions are negligible for small values of m . The functions are alternately positive and negative for large m , so that the modified carrier and side-frequency components, which are always less than the unmodulated carrier, may be

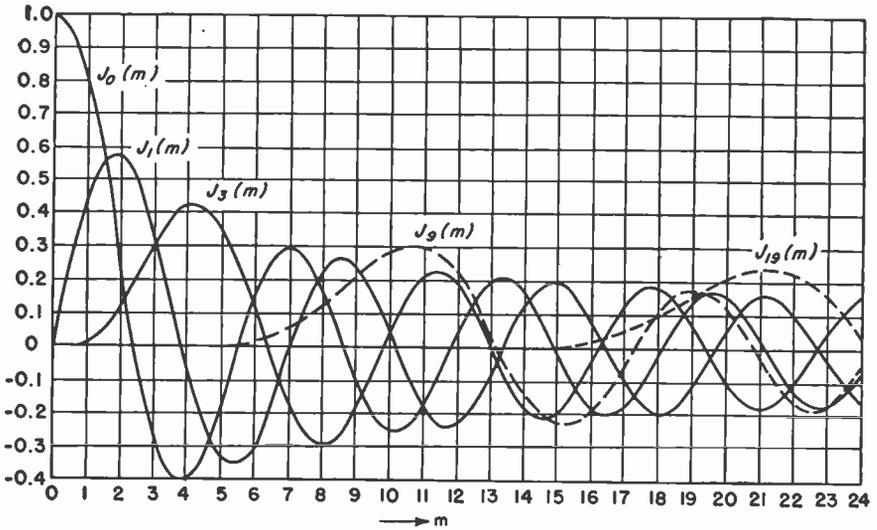


FIG. 8. Examples of Bessel functions of constant order.

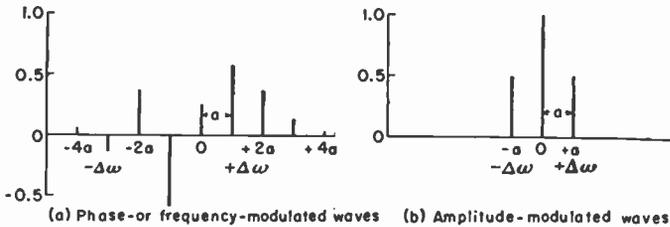


FIG. 9. Component side frequencies.

zero or reversed in phase at certain modulation levels. $J_n(m)$ may be expressed by Eq. (15) when m is less than unity.

$$\left. \begin{aligned}
 J_0(m) &= 1 - \frac{m^2}{4} + \frac{m^4}{64} + \dots \\
 J_1(m) &= \frac{m}{2} - \frac{m^3}{16} + \dots \\
 J_2(m) &= \frac{m^2}{8} - \frac{m^4}{96} + \dots \\
 J_3(m) &= \frac{m^3}{48} + \dots \\
 J_4(m) &< 0.003
 \end{aligned} \right\} \quad (15)$$

Figure 9a illustrates the carrier and side-frequency components for p.m. or f.m. for $m = 2$ for one modulating frequency; Fig. 9b the components for 100 per cent a.m.

for comparison. The odd-order lower side-frequency components have negative signs in p.m. or f.m. to preserve the proper time relations, whereas in a.m. the signs are all positive. The signs of both upper and lower side-frequency components are further reversed when $J_n(m)$ is negative.

The instantaneous phase relations are illustrated in Fig. 10 which shows the constant-amplitude locus of the modulated wave and its phase relations with the carrier wave for $m = 2$. The curve is shown for only the first 180 deg of the modulating cycle since that for the next 180 deg is an image replica of the first about the horizontal axis. The upper and lower side-frequency components, U_n and L_n , are shown when the audio phase at is 0.262 radian or 15 deg. The pairs of components for even values of n combine to produce vectors relative to the carrier

$$V_n = J_n(m) \cos nat \tag{16}$$

in the same phase, while those for odd values of n produce vectors

$$V_n' = jJ_n(m) \sin nat \tag{17}$$

which lead the carrier by 90 deg.

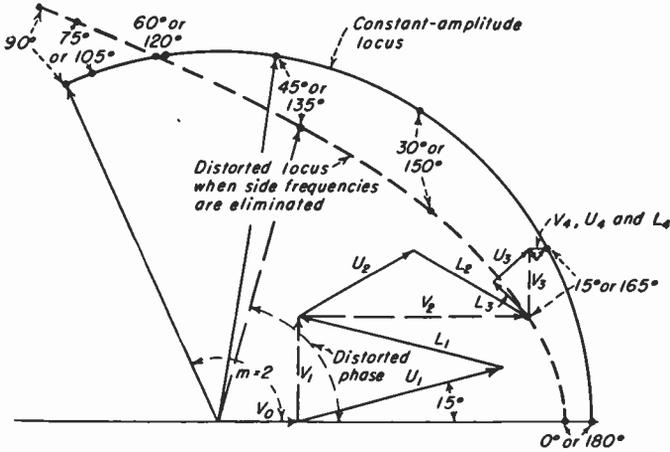


FIG. 10. Loci for $m = 2$.

If any of the side-frequency components are eliminated, as by a filter, the amplitude of the modulated wave is no longer constant and the instantaneous phase no longer represents the original modulation. This is illustrated in Fig. 10. The vector when $at = 45$ deg is shown to be moved from the solid-line to the dashed-line position when the third- and fourth-order side-frequency components are eliminated. The complete locus for this condition is shown by the dashed line.

If the complete wave is shifted in phase by a network having uniform output with uniform time delay, the only effect is to delay the phase of the modulation by a corresponding amount. Nonuniform delay produces distortion.

6. P.M. and F.M. with Several Signal Frequencies. The expansion of the expression for current when several signal frequencies are present contains side-frequency terms which are spaced above and below the carrier frequency by integral multiples of the individual signal-frequencies, and also by the sums and differences of integral multiples of the several signal frequencies. The highest value of n which gives significant Bessel coefficients for the modulation due to any one signal frequency is given by Eq. (14) where the m applies to the same signal component. The frequencies and amplitudes of the first terms of the complete side-frequency spectrum resulting from modulation at two angular frequencies a and b , with modulation

indexes *A* and *B*, are given in Table 1, where *n* and *p* indicate the orders of the Bessel coefficients. The spectrum is more complicated when more than two signal frequencies are present.

There is no upper limit to the degree of modulation which may be applied without distortion by p.m. or by f.m. except as determined by the capability of the equipment

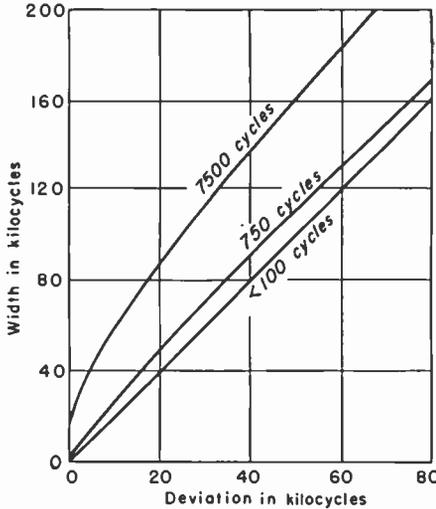


FIG. 11. Band width for single tones.

to operate over the required frequency band. The communication band width required is determined by the side-frequency component of significant amplitude which differs most in frequency from the carrier, since the lack of transmission of any component results in harmonic and intermodulation distortion in the modified wave. The required band differs with the amplitude and frequency of the several modulating components in a complex manner which is indicated partially by the terms in Table 1. The band widths required to transmit all the side bands greater than 1 per cent of unmodulated carrier, when the modulation is due to a single audio tone, are shown in Fig. 11. The data were calculated from n_m as obtained from Eq. (14) multiplied by two times the modulating frequency. The widest band is required when the frequency deviation at high audio rates is a maximum. There are seldom high-pitched

solo notes in music or speech loud enough for full modulation, so while the band required for f-m broadcasting is more than 150 kc, it is not so great as indicated by the upper curve of the figure. It may be of the order of 200 kc, depending on the proportion of different tones in the composite modulating signal.

Table 1

Component	Angular frequency	Relative amplitude*
Carrier.....	ω	$J_0(A)J_0(B)$
Simple side bands of <i>a</i>	$\omega + a$	$J_1(A)J_0(B)$
	$\omega - a$	$-J_1(A)J_0(B)$
	$\omega \pm 2a$	$J_2(A)J_0(B)$
Simple side bands of <i>b</i>	$\omega \pm na$	$(-1)^n J_n(A)J_0(B)$
	$\omega \pm pb$	$(-1)^p J_0(A)J_p(B)$
Combination side frequencies.....	$\omega \pm (a + b)$	$J_1(A)J_1(B)$
	$\omega \pm (a - b)$	$-J_1(A)J_1(B)$
	$\omega \pm (2a + b)$	$J_2(A)J_1(B)$
	$\omega \pm (2a - b)$	$-J_2(A)J_1(B)$
	$\omega + a \pm 2b$	$J_1(A)J_2(B)$
	$\omega - a \pm 2b$	$-J_1(A)J_2(B)$

* To values of *n* and *p* where $J_n(A)J_p(B)$ are negligible.

INTERFERENCE

7. Modulation Due to Interference. All spurious side-frequency components in the acceptance band of a receiver are equivalent to spurious modulation and are amplified and demodulated together with the desired signal. Those spurious com-

ponents which lie within the actual communication band give an output in spite of all possible precautions which may be taken in the design of the receiver. Those components which lie in the adjacent bands may be reduced by increasing the selectivity of the tuned circuits and/or by reducing the cutoff frequency of the audio circuits. When these expedients are carried to an extreme in an a-m receiver, it is at the expense of a reduction in the h-f components of the desired signal, as is often found in low-priced broadcast receivers. In an f-m receiver too great a selectivity leads to harmonic distortion of the desired signal. In general, the response to interference may be analyzed only in connection with the over-all band-pass characteristics of a receiver and differs for a.m. and f.m.

These effects are in addition to the spurious responses which appear in a given receiver due to nonlinearity of the input-output characteristics of vacuum tubes which are covered in other appropriate paragraphs.

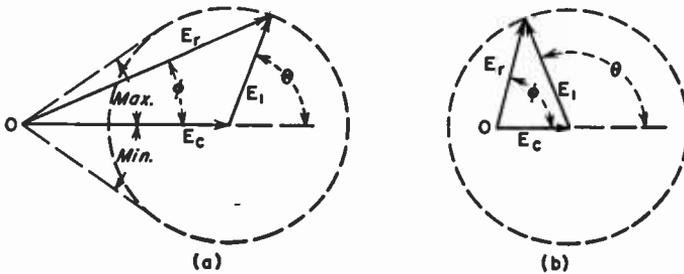


FIG. 12. Combination of single spurious component with carrier.

The sources of interference in the communication band for a particular service are: (1) frequency spectra from undesired radio transmissions, (2) random-noise spectra originating in the antenna and receiver circuits, and (3) spectra from impulse discharges in other electrical equipment. Items (2) and (3) are usually called *random noise* and *impulse noise*, respectively.

8. **Combination of Spurious Components with a Desired Carrier.** A single undesired component E_1 at angular velocity ω_1 combines with a desired unmodulated carrier E_c at angular velocity ω_c as shown by the vector diagrams of Fig. 12. The observer is assumed to be riding with the desired carrier which is shown stationary so that the undesired vector apparently moves at angular velocity $\omega_1 - \omega_c$. The vectors may be located along the reference axis at $t = 0$ without loss of generality so that the phase difference θ is $(\omega_1 - \omega_c)t$. The resultant wave relative to the desired carrier is then

$$\begin{aligned}
 e_r &= E_c \cos \omega_c t + E_1 \cos \omega_1 t \\
 &= E_c \sqrt{1 + p^2 + 2p \cos \theta} \cos \left(\omega_c t + \tan^{-1} \frac{p \sin \theta}{1 + p \cos \theta} \right) \quad (18)
 \end{aligned}$$

where $p = E_1/E_c$.

The peak value varies between $E_c(1 + p)$ and $E_c(1 - p)$. The phase varies between the extremes where the resultant is tangent to the locus when E_1 is less than E_c as shown in Fig. 12a, but continues to rotate between maximum and minimum velocities when E_1 is greater than E_c , as shown in Fig. 12b.

A single component produces the equivalent of both a.m. and f.m. The demodulated beat-note spectrum may be resolved into spurious signal components for both types and includes the fundamental and harmonics of the difference in frequency from the carrier. The harmonics are very pronounced when E_1 and E_c are of nearly equal magnitude, while a nearly pure single-frequency tone is obtained when one is several times the other.

9. Interference from A-m Stations. An interfering a-m wave modifies the vector diagram of Fig. 12a as shown in Fig. 13. The amplitude of the interference varies at the modulating frequency of the interfering wave between the loci E_{1+} and E_{1-} . The degree of the resulting a.m. is $pm/(1 + p)$ or $pm/(1 - p)$, and there is no f-m component, when θ is 0 or 180 deg, respectively, where m is the original degree of modulation of the interfering wave. A lower degree of a.m. is sometimes usefully converted to 100 per cent a.m. in laboratory equipment by making $p = 1/(1 + m)$ and maintaining θ at 180 deg.

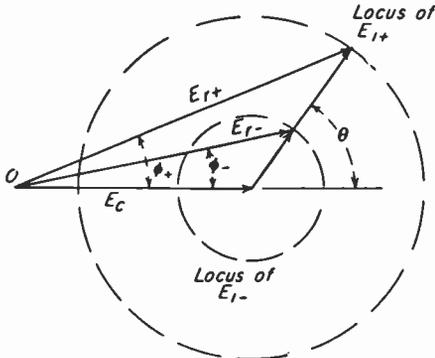


FIG. 13. Interference from a-m wave.

The resulting a.m. consists almost entirely of even harmonics of the interfering modulation and the resulting p.m. (or f.m.) is a maximum when the interfering carrier and the resulting carrier differ in phase by 90 deg. The conversion from a.m. to p.m. is then substantially distortionless if θ is maintained constant and $\phi_+ - \phi_-$ is less than about 50 deg. F.m. may be obtained by making the original a.m. inversely proportional to the signal frequency.

When the interference is an a-m wave on the same channel as a desired transmission, the resulting distortion alternates between a.m. and p.m., and

the demodulated signal contains an audible beat note at the difference of the carrier frequencies. It is heard in receivers designed for either a.m. or f.m. This type of interference may be audible in a-m receivers for levels of interfering carrier as much as 60 db below the desired carrier and is independent of receiver band width.

Synchronized carriers have been transmitted with the same a.m. from antennas separated by 100 miles or more. The value of θ in the resulting wave, in regions where the carriers are of comparable magnitude, then depends on the difference in transmission time, and the pulsations in signal amplitude and distortion are demonstrated readily by listening to a receiver operated in a motorcar as it moves over the road.

The pulsations in amplitude and phase are at a superaudible rate if the carrier of an interfering station is separated by more than one channel from the desired carrier, but the undesired modulation is then detected and heard in the output of an insufficiently selective a-m receiver.

A.m. which is forced intermittently beyond 100 per cent is not only distorted in its own channel but, owing to the discontinuous character of the resulting wave, produces a side-frequency spectrum which is several times the normal width. During swings to overmodulation, the extra side-frequency components then appear in the assigned communication bands of adjacent channels and cause unintelligible monkey chatter therein.

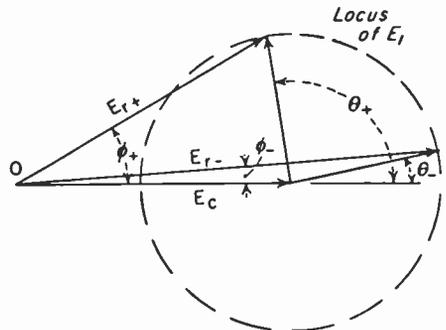


FIG. 14. Interference from p-m or f-m wave of small modulation index.

10. Interference from F-m Stations. An interfering p-m or f-m wave with low m modifies the vector diagram of Fig. 12a as shown in Fig. 14. The resulting a.m. is all even-harmonic distortion of the original p.m. when θ is 0 or 180 deg. The resulting

p.m. is less than the original, when θ is zero, and may be greater than the original and is reversed, when θ is 180 deg. It is accompanied by odd-harmonic distortion in both cases. The p.m. in the resulting wave is a minimum when the interfering and derived carriers are at 90 deg. The a.m. is greatest under these conditions and, when the original m is less than about 0.5 and θ is held constant, is nearly undistorted.

An interfering p-m or f-m wave with low m in the same channel with a desired stronger unmodulated carrier is detected in a receiver designed for any type of modulation with an accompanying beat note and pulsations in amplitude and distortion. The amplitude of such an output is ordinarily so small as to be negligible.

The index m may be hundreds or thousands of radians in f-m broadcasting so that the interfering vector E_1 may rotate several times in alternate directions in each modulating cycle according to the identity $\theta = m \cos at$. The resulting wave has modulation characteristics which are alternately a.m. and f.m. and lead to a beat-note pattern which is repeated approximately once in each cycle of interfering modulation.¹ A limiter, or a nonamplitude responsive frequency detector, is used in most f-m receivers to avoid the effects of a.m., but the frequency variation of the beat note remains and is detected. The amplitude of the unfiltered beat note is proportional to the instantaneous frequency departure of the interfering wave from a desired unmodulated carrier. The beat-note pattern on the screen of an oscilloscope is useful in the laboratory for determining frequency deviation since a calibrated reference generator may be used to determine zero beat. The frequency of the beat note is often above audibility for large portions of the modulating cycle. The peak amplitude of its h-f components is reduced considerably in the output by the deemphasis restoring filter.

The stronger of two f-m broadcast transmissions on the same channel eliminates all intelligible replicas of the original modulation of the weaker when the carrier input ratio is greater than about 6 db. The power in the beat note in the output is no greater than about 0.01 per cent of that for the same modulation on the desired carrier, even when the carriers are equal, because of the reduction in the restoring filter, and is further reduced in proportion to the ratio of the desired to the undesired carrier. The stronger of two wide-band transmissions is, therefore, received with practically no interference from the weaker.

A disturbing and unavoidable type of distortion appears when an f-m transmission is received over two paths that differ slightly in length, as by direct and reflected waves. The instantaneous phase and amplitude of the resulting wave depend critically on the instantaneous phase of the direct and delayed waves as determined by the modulation.² The severe distortion of the signal depends on the modulation index, the signal frequency, and the time difference and relative attenuation over the multiple paths.

11. Random Noise. Components of voltage due to thermal agitation in resistances are equivalent to those derived from unit impulses (time derivatives of step functions) which are random in time (phase). A unit impulse of voltage gives the same analytical results as actual pulses which persist for a time which is much less than the time constants of the circuits involved. It is a convenient mathematical representation of an infinite voltage that exists for an infinitesimal time, wherein the area under the voltage-time curve is unity. The frequency spectrum of a single unit impulse is an infinite number of infinitesimal components of uniform amplitude spaced incrementally in frequency between plus and minus infinity. Since in thermal noise the unit impulses are random in time, the net instantaneous noise voltage has no cyclic character, and the effect is best considered in terms of random-noise power or of effective mean-square voltage.³

The open-circuit mean-square noise voltage in any resistance is

$$e^2 = 4KTR \quad \text{volts}^2/\text{cps of spectrum} \quad (19)$$

¹ Wheeler, H. A., Common-channel Interference between Two Frequency-modulated Signals, *Proc. IRE*, January, 1942, pp. 34-50.

² Corrington, M. S., Frequency-modulation Distortion Caused by Multipath Transmission, *Proc. IRE*, December, 1945, pp. 876-891. Meyers, S. T., Nonlinearity in Frequency-modulation Radio Systems Due to Multipath Propagation, *Proc. IRE*, May, 1946, pp. 256-265.

³ Johnson, J. B., Thermal Agitation of Electricity in Conductors, *Phys. Rev.*, July, 1928, p. 87.

where K = Boltzman's constant, 1.37×10^{-23} joule/°K

R = resistance, ohms

T = absolute temperature

Voltage generated in the radiation resistance of a receiving antenna is applied to the receiver input terminals as an infinite number of spurious components over the acceptance band of the receiver.

Both desired carrier power and noise power are given preferably in terms of the maximum or available power that can be delivered by the antenna to a matched input circuit.¹ This is $V^2/4R$ where V is the rms open-circuit voltage under consideration. The voltage across the input terminals is then half the open-circuit antenna voltage for both carrier and noise. The available power is

$$N = KT \quad \text{watts/cps of spectrum} \quad (20)$$

or approximately 4×10^{-21} watts per cycle per sec for an effective temperature of the antenna radiation resistance of 290°K. Noise calculated on this basis checks with measurements except in isolated cases.

The available noise power at the input is augmented by the contribution of the thermal and shot noise originating within the receiver so that the net noise power is equivalent to a noise input

$$N = FKT \quad \text{watts/cps of band width} \quad (21)$$

where F is the noise factor or noise rating of the receiver. It may be from 1 to 20 depending on the frequency range and the excellence of design.

The noise has components which vary both in frequency and in amplitude, and which, when amplified and demodulated, produce a noise output in either an a-m or f-m receiver. The noise power which reaches the demodulator depends on the effective *energy band width* of the selective circuits and on the gain of the receiver at the center of the pass band, which is the same for carrier and noise when the receiver is properly tuned. The energy band width is the width in cycles of a hypothetical "rectangular" band-pass filter which would pass the same power or mean-square voltage or current as the tuned circuits and is

$$B_e = \left(\frac{1}{A_c} \right)^2 \int_{-\infty}^{\infty} |A(f)|^2 df \quad \text{cps} \quad (22)$$

where $A(f)$ is the voltage transfer ratio or gain at the frequency f , and A_c is its value at the center frequency. In a superheterodyne, B_e is determined mainly by the i-f circuits.

12. Random-noise Power in A-m Receivers. The random-noise components beat with each other in the demodulator in the absence of a carrier to produce components at frequencies equal to their difference, some of which are in the useful output range. A carrier stronger than the noise reduces the spurious noise modulation for the same reasons that the effect of a single interfering component is reduced (see Fig. 12) in either a.m. or f.m.

The signal-to-noise power ratio available at the demodulator in double-side-band a-m receivers is

$$\left(\frac{S}{N} \right)_a = \frac{m^2 A_c^2 C}{N A_c^2} = \frac{m^2 C}{FKTB_e} \quad (23)$$

where C is the available carrier power, m is the degree of desired modulation, and the other terms are as defined above. The signal-to-noise voltage ratio is $mE_o/\sqrt{FKTRB_e}$, where E_o is the available rms carrier input voltage and R is the antenna radiation resistance. The actual interfering noise in the output may be reduced at the expense of loss of h-f signal components by lowering the cut-off frequency of the audio or video amplifier. The over-all signal-to-noise power ratio in double-

¹ Friis, H. T., Noise Figures of Radio Receivers, *Proc. IRE*, July, 1944, pp. 419-422.

side-band a-m receivers is then

$$\left(\frac{S}{N}\right)_a = \frac{m^2 C}{2FKTB_s} \tag{24}$$

where B_s is the over-all energy band width referred to the signal frequency. When neither the selective circuits nor the audio system attenuates the higher frequencies, B_s becomes the highest frequency that can be heard.

The noise power relations as given in the preceding equations are usually sufficient when the desired signal is at an audible frequency, and it is unnecessary to know the envelope of the individual noise pulses as influenced by the selective circuits or to know their time distribution. In devices responsive to instantaneous values, as in cathode-ray indicators used in television and in radar, it is sometimes desirable to know also the crest factor, which is the ratio of the peak voltage to the rms value. The crest factor has been shown by the application of probability to be about 4.0 to 4.5 for random noise, although occasional peaks may greatly exceed this value in the absence of tube overload.¹

The crest factor does not depend on the selectivity, but only on the random distribution with time. The probability that the noise voltage will exceed V at any time is

$$p = \frac{1}{E_n \sqrt{2\pi}} \int_v^\infty e^{-\frac{v^2}{2E_n^2}} dv = \frac{1}{2} \left[1 - \operatorname{erf} \left(\frac{V}{E_n \sqrt{2}} \right) \right] \tag{25}$$

where E_n is the rms noise voltage and *erf* indicates the error integral.

A plot of this curve shows that the noise crests will exceed the rms value 16 per cent of the time, and four times the rms value only 0.0032 per cent of the time. The shape of the envelope for individual noise pulses at the demodulator is approximately the same as obtained with impulse noise which is discussed in Sec. 14, although for random noise the peaks may be expected to vary according to Eq. (25).

13. Random Noise in F-m Receivers. The random noise in an f-m receiver is demodulated by a perfect frequency detector unresponsive to a.m. to produce output voltage components which are proportional in magnitude and in frequency to their spacing from the center frequency. The total noise power produced by the frequency detector, regardless of audibility, when the noise crests are less than the carrier, is²

$$N' = 2A_c^2 FKT \int_0^{B_c/2} \left(\frac{2f}{B_c}\right)^2 df = \frac{A_c^2 FKT B_c}{3} \quad \text{watts} \tag{26}$$

However, in f-m broadcasting the i-f band width is several times the highest audible frequency h as limited by the cutoff frequency of the ear. The audible noise is then obtained by substituting h as the upper limit in Eq. (26), or

$$N = \frac{8A_c^2 FKT h^3}{3B_c^2} \quad \text{watts} \tag{27}$$

where the other factors are as defined previously. In video work h should be the highest frequency to which the output circuit will respond.

The modulation factor in f-m receivers is $(\Delta f)/(\Delta f)_m$ so that the desired output signal power is $A_c^2 C(\Delta f)^2/(\Delta f)_m^2$, where Δf is the deviation and $(\Delta f)_m$ is the maximum system deviation. The output signal-to-noise power ratio is

$$\left(\frac{S}{N}\right)_f = \frac{C(\Delta f)^2}{(\Delta f)_m^2} \cdot \frac{3B_c^2}{8FKTh^3} \tag{28}$$

¹ Landon, V. D., The Distribution of Amplitude with Time in Fluctuation Noise, *Proc. IRE*, February, 1941, pp. 50-55.

² Crosby, M. G., Frequency Modulation Noise Characteristics, *Proc. IRE*, April, 1937, pp. 472-514.

Ordinarily B_c is made substantially equal to $2(\Delta f)_m$ so that for full modulation

$$\left(\frac{S}{N}\right)_f = \frac{3C(\Delta f)_m^2}{2FKTh^2} \quad (29)$$

The total audible noise power is reduced by the action of the deemphasis restoring filter by the ratio

$$\int_0^h \frac{f^2 df}{1 + 4\pi^2 T_0^2 f^2} / \int_0^h f^2 df = \frac{3}{4\pi^2 T_0^2 h^2} \left(1 - \frac{\tan^{-1} 2\pi T_0 h}{2\pi T_0 h}\right) \quad (30)$$

where T_0 is the time constant of the filter. This ratio is 0.0478 when h is taken as 15,000 cycles and T_0 is 75 μ sec and increases the signal-to-noise ratio given by Eq. (29) by 13.2 db.

There is a random-noise improvement threshold for f.m. when the carrier voltage peaks are always greater than the noise crests. This is obtained when

$$\frac{C}{FKT} = 8B_c \cong 16(\Delta f)_m$$

for a crest factor of 4. The noise predominates for lower carrier input ratios, depresses the signal, and smothers it in noise, since for a portion of the time the phase will jump 2π radians as discussed in Sec. 16 for impulse noise.

There is an improvement ratio in the signal-to-noise ratio for f.m. relative to a.m., above the improvement threshold and for full modulation, of

$$\frac{(S/N)_f}{(S/N)_a} = \frac{3(\Delta f)_m^2}{h^2} \quad (31)$$

This does not include the effect of the deemphasis filter or any audio filtering in the output of the a-m receiver.

14. Impulse Noise. Impulse noise, such as ignition interference, is produced by transient disturbances separated in time by quiescent intervals, and shock excites neighboring conductors which reradiate at frequencies depending on their dimensions and separation from ground. The radiated components usually extend over an extreme frequency range and, within the pass band of the receiver, are of approximately equal amplitude. They differ from those due to random noise in that they are all in phase or are delayed by a uniform time from a reference phase.

The "spike" of impulse noise in volt-seconds at the receiver input may be represented by an amplitude factor multiplied by the unit impulse. The noise power depends on the repetition rate so that the best evaluation is in terms of peak voltage response for each individual impulse.

The response of the high- Q tuned circuits of a receiver to the spike function may be shown by Fourier analysis or operational calculus to be a delayed and broadened a-m pulse on a carrier at the center frequency of the tuned circuits. The small amount of phase modulation which accompanies the pulse may be neglected for all practical purposes when the circuit Q s are of the usual order of magnitude.

The shape of the noise pulse at the demodulator depends on the number of tuned circuits and their degree of coupling. The envelope is broadened, rounded, and delayed, and its peak value is limited. The delay is of the same order of magnitude as the pulse width.

The shape and amplitude of the noise envelope in the absence of a desired carrier may be approximated from the derived Fourier expressions relative to the center of the pulse in time, and relative to the response at the center of the pass band in frequency. These factors are related to the average band width. Thus,

$$E_1 = s \int_{-\infty}^{\infty} A(f) df \quad \text{volts} \quad (32)$$

$$sA_c = \int_{-\infty}^{\infty} E(t) dt \quad \text{volt-sec} \quad (33)$$

$$B_a = \frac{1}{A_c} \int_{-\infty}^{\infty} A(f) df \quad \text{cps} \tag{34}$$

$$t_a = \frac{1}{E_1} \int_{-\infty}^{\infty} E(t) dt \quad \text{sec} \tag{35}$$

$$B_a = \frac{1}{t_a} = \frac{E_1}{sA_c} \quad \text{cps} \tag{36}$$

where $A(f)$ is the voltage transfer ratio or gain at frequency f and A_c is its value at the center frequency; $E(t)$ is the voltage envelope at the demodulator at time t , and E_1 is its peak value; s is the volt-second available input noise spike; B_a is the average band width of a hypothetical "rectangular" filter which will pass the same average voltage or current as the actual circuits; and t_a is the width of a rectangular pulse having the same volt-second area.

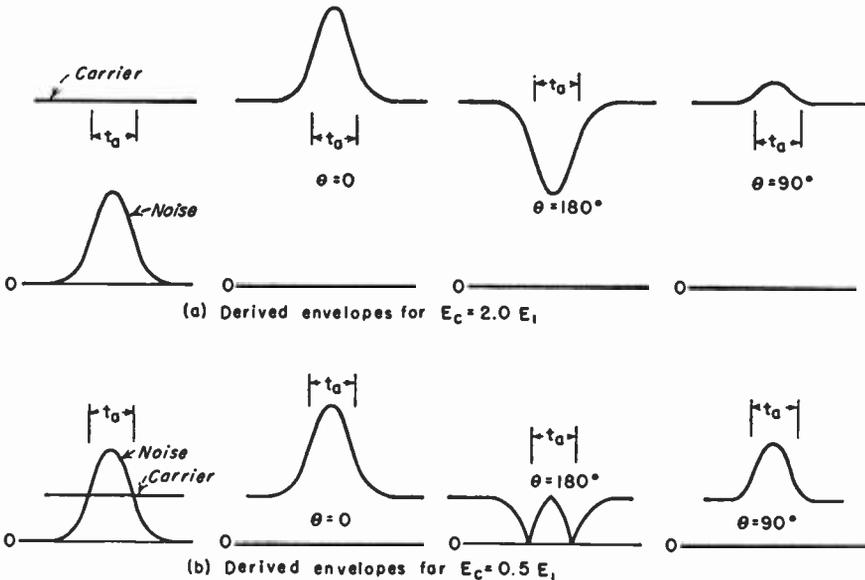


FIG. 15. Illustrative impulse-noise envelope for $E(t) = E_1 \epsilon^{-\pi B_a^2 t^2}$.

The average band width B_a is greater than the energy band width B_e [Eq. (22)], and neither is the same as the band width at half power, sometimes used in defining filters.

In many cases the noise envelope may be approximated by the probability curve,

$$E(t) = sA_c B_a \epsilon^{-\pi B_a^2 t^2} \quad \text{volts} \tag{37}$$

for the corresponding transfer characteristic relative to the center frequency

$$A(f) = A_c \epsilon^{-\pi f^2 / B_a^2} \quad \text{volt-sec} \tag{38}$$

15. Impulse Noise in A-m Receivers. A noise pulse modifies the received wave depending on the relative phase of the desired carrier and transient waves in a manner similar to that shown in Fig. 13 except that the amplitude of the applied-noise envelope starts from zero, reaches E_1 , and returns to zero, during the noise pulse. Illustrative resulting envelopes are shown in Fig. 15 when the carrier is substantially at the center frequency of the tuned circuits. One value of θ is just as likely to occur as any other so that the average envelope is less than the peak. The inward and outward

noise modulations are substantially equal when the noise peak is smaller than the carrier. The outward modulation is greater than the inward when the noise peak is large, since the effective inward modulation can never exceed the carrier.

The peak signal-to-noise voltage ratio at the demodulator is

$$\left(\frac{s}{n}\right)_a = \frac{mE_c}{sB_a} \quad (39)$$

or, if limited by audibility,

$$\left(\frac{s}{n}\right)_a = \frac{mE_c}{2sh} \quad (40)$$

where the terms are as defined previously. Average signal-to-noise ratios for noise peaks less than the carrier are approximately $2/\pi$ times the above values. Average signal-to-noise ratios for large noise are less definite, as illustrated by Fig. 15.

The signal-to-noise ratio in the output depends on whether the detector is linear or square-law, and on the band-pass characteristics of the audio or video circuits.

The situation is substantially the same when the carrier frequency differs slightly from the center frequency of the tuned circuits except that θ progresses through a definite angle during the noise pulse, thereby modifying slightly the shape of the derived envelope.

16. Impulse Noise in F-m Receivers. The phase of an unmodulated carrier is influenced by a single noise pulse as illustrated in Fig. 13 except that the noise wave starts from, and returns to, zero during the period of the pulse. The maximum phase shift due to a pulse, corresponding to the most unfavorable value of θ , for pulses smaller than the carrier, is approximately $\tan^{-1} E_i/E_c$, and the peak noise voltage from an ideal frequency detector, regardless of audibility, is

$$E_i' = 2sA_c \int_0^{B_a/2} \left(\frac{2f}{B_a}\right) df = \frac{sA_c B_a}{2} \quad (41)$$

where the other terms are as defined previously.

The audible peak voltage in f-m broadcasting as limited by the highest audible frequency h , is

$$E_i = \frac{2sA_c h^2}{B_a} \quad (42)$$

The peak signal-to-noise voltage ratio is

$$\left(\frac{s}{n}\right)_f = \frac{E_c B_a}{2sh^2} \quad (43)$$

This does not include the effect of the deemphasis restoring filter which reduces the audible noise peak by 8.5 db when h is taken as 15,000 cycles and T_0 is 75 μ sec. The average signal-to-noise ratio for small noise is substantially $2/\pi$ times the above value since the angle θ at the pulse peak is accidental.

The conditions are modified somewhat during deviation by desired modulation or when the signal is not adjusted to the center of the pass band, but the average before detection is substantially the same.

The impulse-noise signal-to-noise voltage-improvement ratio for full modulation for f.m. relative to a.m. above the *improvement threshold* where the carrier peaks exceed the noise peaks is

$$\frac{(s/n)_f}{(s/n)_a} = \frac{B_a}{h} \quad (44)$$

This does not include the effect of the deemphasis filter or of any audio cutoff filter in the a-m receiver.

The noise output becomes less dependent on the carrier when the noise peaks are nearly equal to the carrier since the maximum phase shift is then limited to ± 90 deg.

The conditions are approximately the same when the noise peaks exceed the carrier except during that portion of the time when the phase of the desired wave, as deviated by modulation, passes through approximate phase opposition to the noise carrier during the period of the pulse.¹ The desired phase then loses 2π radians within the period of the applied noise pulse, and the output of the frequency detector is a unidirectional pulse of voltage whose amplitude is proportional to the average band width of the tuned circuits but which is independent of the magnitude of the noise pulse that produces it. It induces a transient in the deemphasis circuit which momentarily reduces the desired signal output by an audio transient whose wave form is that of an RC circuit having the same time constant. The resulting noise is a loud low-pitched "pop" in contrast to the faint "clicks," which contain mainly h-f components and characterize impulse noise in the absence of pops. The chances of producing a noise pop increase with mistuning and with instantaneous modulation.

A frequency detector which is balanced, *i.e.*, one giving no output at the center frequency of the receiver, is of particular value in reducing the output due to impulse noise. The rapid successive positive and negative frequency excursions that accompany each click type of pulse then largely neutralize each other before the output circuit can respond. It is also important that the selective circuits be symmetrical and the frequency-detector characteristic be skew-symmetrical with respect to the center frequency. Finally, the receiver should be accurately tuned to the desired signal.² The noise output is then only that discussed for the ideal frequency detector.

AMPLITUDE-MODULATED WAVES IN NONLINEAR CIRCUITS

17. Modulation, Frequency Conversion, and Detection. Essentially the same classes of nonlinear devices are used for the modulation, frequency conversion, or detection of a-m waves. In each process the waves to be *combined* or *resolved* are applied to circuit elements which have asymmetrical EI characteristics. These may be series characteristics, as in dry rectifiers or diodes, or may be mutual characteristics, as in multielectrode vacuum tubes.

The output circuit of a modulator is arranged to transmit the carrier and its side bands; that of a frequency converter, the i-f carrier and its side bands; and that of a detector, the components at the frequency of the original signal. The components in the voltage developed by the output current at other frequencies are eliminated by proper filtering.

Intermodulation is the production of new components having frequencies corresponding to undesired sums and differences of the fundamental and harmonic frequencies of the components of the applied waves.

Cross modulation is a type of intermodulation in which the carrier of the desired output signal is modulated by an undesired signal.

Modulation distortion is a change in the character of modulation either in a change in the percentage of modulation or in the production of harmonics of the modulating signal due to intermodulation.

Spurious modulation components may be predicted by substituting the desired and interfering input signals in the expressions for plate current, expanding, and collecting the terms at the frequencies in question.

18. Two-terminal Rectifiers. Units having asymmetrical IE characteristics and, therefore, having inherent rectifying properties, are suitable as detectors or demodulators. The earliest two-terminal signal rectifier was the delicate "cat-whisker" contact of a fine wire with a sensitive spot on the surface of a natural galena or iron pyrites crystal. Its ratio of forward-to-backward conductance was uncertain, owing to changes in contact pressure and to surface oxidation. This early crystal has been made obsolete by the high-vacuum diode tube which was highly satisfactory until

¹Smith, D. B., and W. E. Bradley, The Theory of Impulse Noise in Ideal Frequency Modulation Receivers, *Proc. IRE*, October, 1946, pp. 743-751.

²Landon, V. D., Impulse Noise in F-M Reception, *Electronics*, 14, 26-30, 73-76, February, 1941.

signal rectifiers or converters were required for centimeter waves, where interelectrode capacitance and the transit time for electrons to pass from cathode to anode badly limited diode performance. Both of these factors were reduced by using small, closely spaced electrodes, making operation possible up to a few hundred megacycles.

The recently developed silicon crystals with fine tungsten-wire cathodes are by far the most effective converters and video detectors for the centimeter waves. In these, the carefully shaped point of a fine tungsten spring bears against a polished surface of the crystal, and the whole assembly is adjusted permanently and sealed in a small cartridge. The chief shortcoming of silicon crystals is their susceptibility to burnout. Instantaneous overload or even static discharges through the crystal from the fingers in handling ruin them. A similar construction with germanium instead of silicon makes a more rugged high-current rectifier particularly suited for frequencies not exceeding a few hundred megacycles.

Copper-oxide rectifiers¹ are satisfactory at low carrier frequencies, but they have high self-capacitance which prevents their use at frequencies above a few megacycles. Selenium rectifiers, recently developed for power-supply purposes, require time to block in the reverse direction and are not used as signal detectors.

All rectifiers have instantaneous current-voltage characteristics which are difficult to express over a wide range of applied voltages. However, these curves in the forward or highly conducting direction, are characterized by a slope ($\delta i/\delta e$) which increases with increasing applied voltage. For vacuum diodes

$$i = h(e + e_0)^{3/2} \quad (45)$$

where e is the instantaneous applied voltage; h increases with cathode area; and e_0 increases with cathode temperature. For crystal diodes, over a limited range,

$$i = ke^n \quad (46)$$

where k and n are both functions of e and of temperature. The slope of the curve may vary from a constant at low voltage and high temperature to values proportional to the cube of the voltage at high voltage and low temperature. Typical characteristics for a silicon diode at room temperature and about 1 volt are $i = 0.03e^{1.5}$; for a germanium diode, $i = 0.01e^{1.7}$.

At zero applied voltage, crystals and copper-oxide rectifiers pass zero current, while vacuum-tube diodes have an internal voltage (contact potential) and circulate current depending on the connected circuit resistance. Crystals and copper-oxide rectifiers have high finite resistances for reverse applied voltage, whereas vacuum diodes cut off completely for reverse voltages of more than a fraction of 1 volt. The resistance of crystals in the backward direction may be of the order of megohms at a few volts. Germanium crystals exhibit negative resistance when the backward current exceeds a few milliamperes as produced by the application of reversed voltage of the order of 100 volts. In the negative-voltage region before cutoff, vacuum diodes have a characteristic expressed approximately by

$$i = he_0^{3/2} e^{e/e_1} \quad (47)$$

where e_1 is nearly independent of operating conditions and the other factors are as defined previously.

19. Input to a Single Grid. The grid-plate characteristic of a vacuum tube in which the plate current is substantially independent of the load may be represented by the power series

$$i = A_0 + A_1e + A_2e^2 + A_3e^3 + \dots \quad (48)$$

where the A s are coefficients determined by test and e is the instantaneous input voltage. Specifically, the coefficients are

¹ Caruthers, R. S., Copper-oxide Modulators in Carrier Telephone Systems, *Trans. AIEE*, **58**, 253-265, June, 1939.

$$A_n = \frac{1}{n!} \frac{\partial^n i}{\partial e^n} \tag{49}$$

at the steady value of e which is maintained by bias voltage.

The power series Eq. (48) is useful quantitatively as well as qualitatively in class A tubes which do not draw grid current and which are not worked to plate-current cutoff. In class B and class C services the power-series equation may require too many terms for the extended range and may converge too slowly to be of use quantitatively, but it is always of value in determining the frequency range of the possible output components. The term $A_2 e^2$ provides the largest part of the useful modulation output, whereas the term $A_1 e$ provides the useful output in amplification. When the higher order terms are absent, as at low input levels in class A tubes, the useful output may be calculated accurately.

In triodes (or when the output impedance is appreciable with respect to the plate impedance), the plate current depends on the plate voltage as well as on the grid voltage. The voltages developed by the plate current in the plate load are reimpressed on the plate, thereby reducing the useful output. To a first approximation, neglecting the higher order terms, the output current from Eq. (48) is reduced by the factor $r_p / (r_p + Z)$ where r_p is the plate resistance of the tube and Z is the load impedance at the frequency of the desired output. If the load impedance is nonuniform over the band, the correction may be applied separately to each of the output-frequency components.

20. Input to Two Electrodes. The plate current of a vacuum tube, in which the plate current is substantially independent of the load and which is controlled by the potentials on more than one electrode, may be expressed by the double-power series

$$i = A_0 + A_1 e_1 + A_2 e_1^2 + A_3 e_1^3 + \dots + e_2 (B_0 + B_1 e_1 + B_2 e_1^2 + \dots) + e_2^2 (C_0 + C_1 e_1 + C_2 e_1^2 + \dots) \tag{50}$$

in which the applied voltages are e_1 and e_2 and in which the coefficients are determined by test.

$$\begin{aligned} A_1 &= \frac{\partial i}{\partial e_1} = g_1 \\ A_2 &= \frac{1}{2} \frac{\partial^2 i}{\partial e_1^2} = \frac{1}{2} \frac{\partial g_1}{\partial e_1} \\ B_0 &= \frac{\partial i}{\partial e_2} \\ C_0 &= \frac{1}{2} \frac{\partial^2 i}{\partial e_2^2} \\ B_1 &= \frac{\partial^2 i}{\partial e_1 \partial e_2} = \frac{\partial g_1}{\partial e_2}, \text{ etc.} \end{aligned} \tag{51}$$

at the steady values of e_1 and e_2 , which are maintained by bias voltages.

The double-power series Eq. (50) may be used qualitatively for an estimate of the possible frequency components present in the output, even if it converges too slowly for quantitative results.

The term $B_1 e_1 e_2$ provides useful modulation output. When the higher order terms are small, as at low input levels in class A operation, the useful output may be calculated accurately.

21. Linearity of Output. While the power-series equations are extremely useful in class A calculations and in showing what distortion components may be present under less favorable circumstances, a direct indication of the linearity of the desired output in terms of the variable input is more often used. Tests of vacuum-tube modulators and detectors may be conducted at any convenient frequency, *e.g.*, 60 cycles, if the circuit impedances at the desired operating frequency are duplicated

at the test frequency. The linearity of the dynamic characteristic of the controlled current is a direct indication of the linearity of the desired output.

For modulators the output current (or the voltage developed in the load by it) at the test frequency is plotted against steps in the input voltage corresponding to its variation by the modulating voltage. For detectors, the d-c output is plotted against steps in the rms or peak value of the test input voltage, corresponding to the modulated input voltage for different parts of the modulating cycle.

The output curves which are the most nearly linear over a wide range of the independent variable, taken for different load resistances, bias voltages, etc., indicate the operating condition which will accommodate high percentages of modulation with the least distortion.

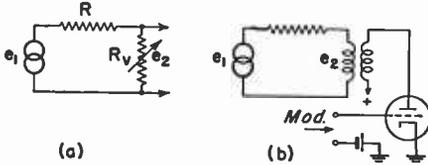


FIG. 16. Method of absorption modulation.

in accordance with some function of the modulating voltage. In Fig. 16a suppose that R_v is a resistance which includes the load and which is varied linearly with the modulation or

$$R_v = R_0(1 + M) \tag{52}$$

where M = instantaneous value of the modulation
The output voltage e_2 is

$$e_2 = \frac{e_1 R_v}{R + R_v} \tag{53}$$

when e_1 = applied voltage
 R = resistance of the source

The ratio of e_2/e_1 is plotted against R_v/R in Fig. 17, which indicates that reasonably linear operation is obtained over a small portion of the curve when R_v is small compared to R . The dashed curve is the output voltage across R as a load when R_v is used as a variable series resistor. Physical resistances can be varied over a limited range, say three to one. The usable portion of the curve of Fig. 17 is then limited to the section marked ab . The efficiency and the effective degree of linear modulation in the output are low in any case.

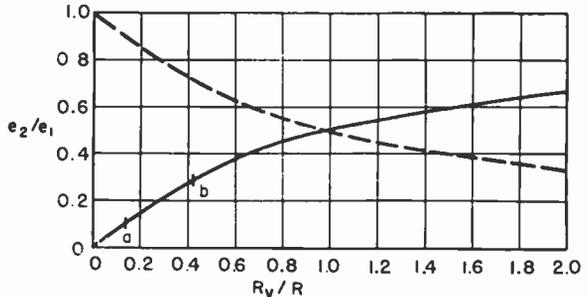


FIG. 17. Linearity in absorption modulation.

The plate resistance of a vacuum tube as controlled by the modulating voltage applied to its grid may be used as the variable resistor for absorption modulation in parallel with the output load as shown in Fig. 16b. Since the plate resistance is not a linear function of the grid voltage, the over-all linearity may then be improved somewhat by working on a portion of the tube characteristic which tends to cancel the required curvature indicated in Fig. 17. Tests are then made for linearity of load voltage vs. grid voltage.

The plate resistance required of the tube is the inverse of the resistance R_v when a quarter-wave transmission line or its filter equivalent is interposed between the tube

AMPLITUDE MODULATORS

22. Absorption Modulators. Absorption modulation is obtained by varying a resistance either in series with or in parallel with the load in accordance with some function of the modulating voltage.

In Fig. 16a suppose that R_v is a resistance which includes the load and which is varied linearly with the modulation or

and the load. The required plate resistance R_p is then

$$R_p = \frac{Z_0^2}{R_v} \tag{54}$$

where Z_0 = image impedance of the line

Absorption modulation has been used to supplement other modulating methods over portions of the modulating cycle.

23. Grid Modulators. Grid modulators operate with carrier and signal voltages applied to the same or separate grids as illustrated in Fig. 18a or 18b. The plate

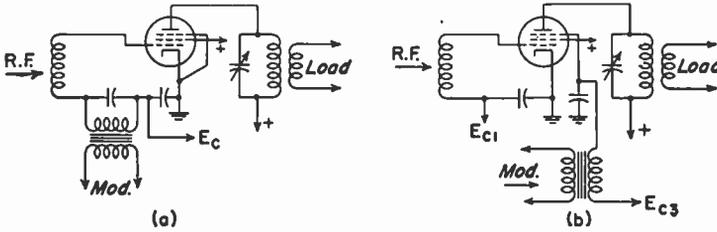


FIG. 18. Types of grid modulators.

current may be calculated by Eq. (48) or (50) for low levels when the plate current is not swung to cutoff. The action is illustrated in Fig. 19 for the connections of Fig. 18a and for a square-law tube. Curve *a* shows the input signal and carrier voltages superimposed, curve *b* the instantaneous plate current, and curve *c* the modulated output voltage with the l-f components filtered out.

In this case, when the applied voltage about the operating point E_c' is

$$e = E \cos \omega t + S_1 \cos a_1 t + S_2 \cos a_2 t + \dots \tag{55}$$

where S_1 and S_2 are the signal amplitudes at frequencies $a_1/2\pi$ and $a_2/2\pi$, etc., the useful output current is

$$i = E(A_1 + 2A_2 S_1 \cos a_1 t + 2A_2 S_2 \cos a_2 t + \dots) \cos \omega t \tag{56}$$

This may be written simply

$$i = E(A_1 + 2A_2 M) \cos \omega t \tag{57}$$

where M indicates the instantaneous applied modulating signal. The product $M \cos \omega t$, when expanded, produces all the pairs of side frequencies required for the modulated wave. There are no spurious modulation components. However, this mode of operation does not realize fully the power capability of the tube, and the modulation cannot approach unity.

A grid modulator is operated as a class C carrier-frequency amplifier for higher plate efficiency. A tube with a linear grid-plate characteristic is suitable. The bias is adjusted to about twice the value required for plate-current cutoff, and the carrier input voltage is adjusted until the peaks reach halfway between saturation and plate-current cutoff. The superimposed modulating voltage at signal frequency causes

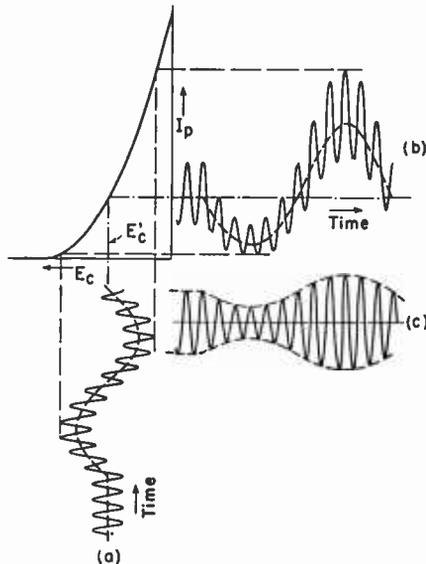


FIG. 19. Low-level grid modulation.

the positive peaks to vary widely in value as shown by curve *a* in Fig. 20. Curve *b* shows the instantaneous plate current and curve *c* the useful modulated output voltage with the 1-f and carrier-harmonic components filtered out.

Linearity may be tested by the method of Sec. 21 and observing the output at the test frequency for a range of bias voltages. The exact bias setting is then at the center of the linear portion of the test curve.

Grid modulators have the advantage of requiring small signal input power, particularly when the tubes are not driven to grid current, but have limited ranges of

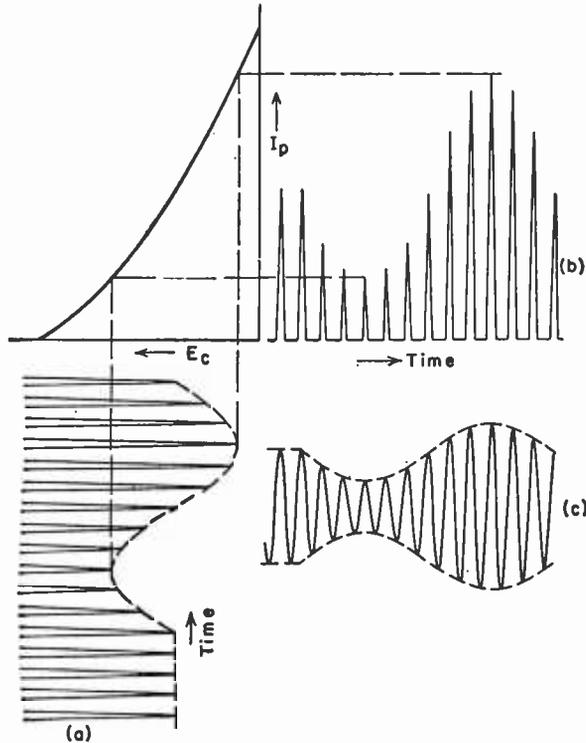


FIG. 20. Class C grid modulation.

linear modulation and plate efficiencies of only 20 to 30 per cent. They are used ordinarily at low power levels.

Grid modulators are used for television signals since it is difficult to obtain reasonable operation with high-level plate modulation over the required wide band of television modulation frequencies. Triodes may be used in grid modulators if neutralized to prevent h-f feedback. The output is then reduced as explained in Sec. 17. The voltage, current, and power in the plate circuit have the following approximate relations for sine-wave modulation:

$$\begin{aligned}
 E_{\max} &= E_c(1 + M) \\
 W_{\max} &= W_c(1 + M)^2 \\
 i_{p\max} &= I_c(1 + M) \\
 W_{av} &= W_c \left(1 + \frac{m^2}{2} \right)
 \end{aligned} \tag{58}$$

where the subscript *c* indicates the conditions for the carrier alone.

An extra tube which is made to operate by adjustment of its grid and plate potentials over a portion of the modulation cycle to supplement the first modulator tube is sometimes used to obtain higher efficiency and at the same time preserve or even improve linearity of modulation.¹

24. Balanced Modulators. When carrier voltage is applied in phase and modulating voltage is applied in push-pull to the grids of two modulator tubes, the carrier is

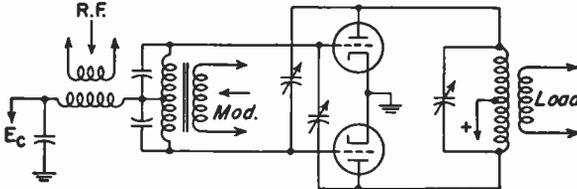


FIG. 21. Circuit for balanced modulator.

balanced out in a push-pull output load. The circuit shown in Fig. 21 with two neutralized triodes is typical.

For exact balance of tubes and transformers, and over the range of signals for which the modulation characteristic is linear, the useful modulated output current from the two plates is

$$i_1 = E_c(A_1 + 2A_2e_m) \cos \omega t \tag{59}$$

and

$$i_2 = E_c(A_1 - 2A_2e_m) \cos \omega t \tag{60}$$

where E_c = maximum carrier voltage

e_m = instantaneous modulating voltage

The effective input current to the tank circuit is

$$i = 4A_2E_c e_m \cos \omega t \tag{61}$$

which contains only the side bands. For a single modulating frequency $a/2\pi$, this reduces to

$$i = 2A_2E_c e_m [\cos (\omega + a)t + \cos (\omega - a)t] \tag{62}$$

The voltage developed in the output circuit is

$$e = 2A_2E_c e_m Z [\cos (\omega + a)t + \cos (\omega - a)t] \tag{63}$$

where Z = load transfer impedance

This arrangement is used in suppressed-carrier transmission systems. It has the advantage of balancing out any even-harmonic distortion due to departure of the modulation characteristic from linearity, not considered in the above equations.

When the modulating input voltages to the two grids, the tube coefficients, and the effective load transfer impedances for the two tubes are unequal, the net tank circuit voltage is

$$e = E_c[A_1'Z' - A_1''Z'' + 2(A_2'e_m'Z' + A_2''e_m''Z'')] \cos at \cos \omega t \tag{64}$$

where the ' and '' values are for the first and second tubes, respectively. Some of the carrier remains when the balance is not perfect.

25. Plate Modulators. The constant-current plate modulator utilizes an a-f choke coil in the circuit which supplies plate power to r-f and a-f amplifier tubes, as shown in Fig. 22. The total plate current remains constant by virtue of the induct-

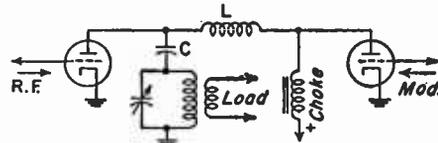


FIG. 22. Circuit for plate modulation.

¹ Terman, F. E., and J. R. Woodyard, A High-efficiency Grid-modulated Amplifier, *Proc. IRE*, **26**, 929-945, August, 1938.

ance of the choke. The instantaneous audio plate voltage is added to the plate supply voltage and over the audio cycle changes the latter to $E_b(1 + M)$. The r-f inductance L prevents the loss of h-f power in the audio tube, and C prevents the short circuit of the audio plate voltage.

The a-f tube, frequently called the *modulator*, supplies modulating power, but the actual modulation occurs in the plate circuit of the r-f tube.

The plate current of the a-f tube cannot be reduced to zero during the modulation cycle without introducing audio distortion. It is necessary, therefore, to operate the audio tube with a higher zero-signal plate current than the radio tube in order to reach unity modulation without audio distortion. This is done by applying a higher plate voltage to the audio tube, either by using a by-passed resistance in the plate circuit

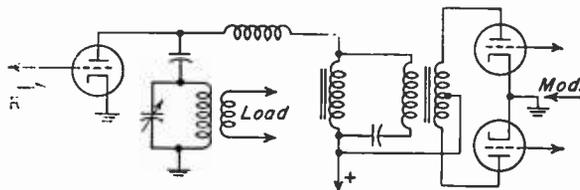


FIG. 23. Transformer-coupled plate modulator.

of the r-f tube or by supplying the audio tube (or tubes) through transformer coupling as illustrated in Fig. 23. A further improvement is indicated in Fig. 23, since saturation of the transformer core is prevented by eliminating the d-c magnetizing component and since even-order audio harmonics are canceled by the push-pull arrangement. The efficiency of the system is increased by operating the audio tubes in push-pull class B. In transformer coupling the total d.c. is no longer constant but varies with the modulation.

The voltage, current, and power in the plate circuit have the following relations for sine-wave modulation:

Tube and circuit voltage,

$$E_{\max} = (1 + m)E_b$$

R-f input power,

$$W_{\max} = (1 + m)^2 E_b I_b$$

Average r-f input power,

$$W_{av} = \left(1 + \frac{m^2}{2}\right) E_b I_b$$

Average output power,

$$W_o = \eta \left(1 + \frac{m^2}{2}\right) E_b I_b$$

Average audio input power,

$$W_a = \frac{m^2}{2} E_b I_b$$

R-f plate loss,

$$W_p = (1 - \eta) \left(1 + \frac{m^2}{2}\right) E_b I_b \quad (65)$$

where E_b and I_b = d-c supply voltage and current

m = degree of modulation

η = plate efficiency of radio tube

The plate efficiency is high (often 0.7 to 0.8), and the chief disadvantage of plate modulation is the large amount of audio power which must be supplied.

The radio tube is operated as a class C amplifier with bias at approximately two times cutoff. Since the plate losses are 50 per cent higher with full modulation

than for unmodulated output, the tube must be used at two-thirds its rated power. Low- μ triodes are suitable and ensure low plate and grid voltages.

Grid-leak bias helps in obtaining linearity up to complete modulation. Linearity may be checked by direct adjustment or by test at 60 cycles with a proper plate load for a range of plate voltage from 0 to $2E_b$.

26. Modulated Oscillator. Plate modulation was originally applied directly to the oscillator tube and circuit. Practically full modulation may be obtained with excellent linearity, but the arrangement has the disadvantage of introducing f.m. The frequency of the oscillator varies with the plate voltage, and, since in plate modulation this varies between $2E_b$ and 0 during full modulation, the oscillator frequency deviates from its mean value with the modulating signal.

The same circuits are used between the two tubes, and the same voltage, current, and power relations hold as with a plate-modulated amplifier. Linearity is obtained by adjusting the value of the grid leak.

Modulated oscillators are now considered suitable only for test equipment in which the f.m. is not objectionable.

27. Copper-oxide Modulators. Copper-oxide rectifiers are applicable in bridge modulators and are used widely in carrier-current telephony. They function as carrier-operated switches for opening, shorting, or reversing the elements carrying the modulating currents. For such purposes they are the familiar dry rectifiers for 60-cycle power in miniature. Each unit consists of a small copper disk, one side of which is oxidized and placed in contact with a soft metal. Electrons flow most easily from the copper to the soft metal. The resistance of a $\frac{3}{16}$ -in. unit at low frequencies is of the order of 100 ohms in the forward direction and from 10^4 to 10^6 ohms in the reverse direction. The current-carrying capacity is roughly proportional to the area of the disk, which may be $\frac{1}{16}$ in. or more in diameter. The voltage rating is increased by placing several disks in series. Bridge units are usually hermetically sealed with only the terminals brought out.

Copper-oxide rectifiers are not suitable for use at frequencies much above 1 Mc, except at low impedance levels on account of inherently large shunt capacitance. They are compact in size and eliminate the heater connections necessary in similar circuits using vacuum tubes. They maintain a satisfactory balance in carrier-suppression circuits using balanced modulators.

The power-series current equation for a copper-oxide unit converges slowly, and its characteristics are expressed more easily quantitatively in terms of resistance for different applied voltages. For voltages in the reverse direction and for less than 0.02 volt in the forward direction, the resistance is high and substantially constant. For forward voltages between 0.02 and 0.6 volt, the resistance is approximately

$$r = r_0 e^{-ks} \tag{66}$$

where k is a constant which may be as great as 18. For forward voltages larger than 0.6 volt, the resistance is low and nearly constant.

Representative circuits using copper-oxide units are shown in Figs. 24 and 25. In these figures, f_c and f_s indicate voltage sources at the carrier and signal frequencies, respectively. The impedances of the carrier source, signal source, and load are Z_c , Z_s , and Z_L , respectively. The forward-conducting direction of the units is shown by the arrows.

Since the carrier and signal voltages are applied to conjugate terminals of the rectifier bridges, the resistance effects are balanced. In Fig. 24 the output signal is short-

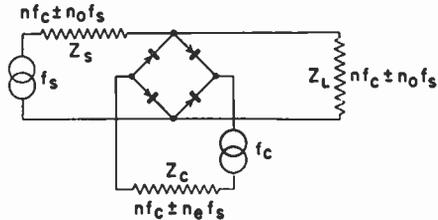


FIG. 24. Bridge modulator using nontube rectifiers.

circuited for one polarity of the carrier cycle. In Fig. 25 the effective connections between the signal source and the load are reversed as the polarity of the carrier changes. This arrangement is called a *double-balanced* or *ring modulator*.

The frequencies of the current components produced in the individual rectifier units are determined qualitatively by an expansion of each of the terms of a power-series equation for the current. Current components of frequencies equal to the sums and differences of the integral multiples of the carrier and signal frequencies appear in

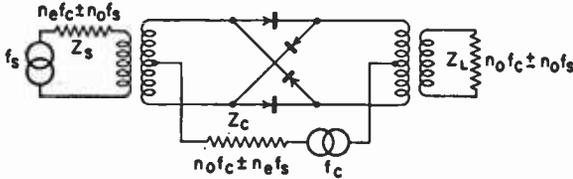


FIG. 25. Double-balanced or ring modulator.

each unit in the forward direction. These combine additively or differentially in the connected circuits depending on the polarity. In flowing through the circuit impedance these current components produce voltages of the same frequency which are reimpressed upon the rectifier units. The final result may be obtained quantitatively only by a series of approximations. The frequencies of the components appearing in the circuit impedances are indicated in Figs. 24 and 25, where n is any whole number or zero, n_o is any odd number, and n_e is any even number or zero.

The output impedances are designed as filters to eliminate voltages at frequencies involving undesired multiples of f_c . The useful output is at a frequency $f_c \pm f_s$, and, when double modulation is used, the carrier frequency f_c is eliminated.

By making the carrier voltage large in comparison with the signal voltage, the terms involving multiples of f_s may be reduced satisfactorily in magnitude. The units are operated with about 0.5 volt carrier across each disk in the forward direction. The optimum impedance in the signal and output circuits is

$$Z = \sqrt{R_f R_r} \tag{67}$$

where R_f and R_r are the resistances in the forward and reverse directions, respectively. The loss in conversion is then only 6 to 8 db.

Copper-oxide bridge modulators differ from van der Bijl vacuum-tube balanced modulators in that they transmit in either direction. They function equally well as modulators or demodulators.

28. Crystal-diode Modulators. The same circuits may be used for germanium crystal diodes as for copper-oxide rectifiers, although at the present time information is not available as to their stability of balance in bridge circuits. They have low self-capacitance, and single units are operable to frequencies of several hundred megacycles. The forward current of germanium crystals must not exceed about 200 ma, transient, or 20 ma, average. The maximum reverse voltage should not exceed about 50 volts.

PULSE MODULATORS

29. Pulsed Oscillators. Pure pulse-modulated u-h-f waves are produced by directly controlling a power oscillator, coupled through a transmission line or wave guide to the transmitting antenna, since modulated power amplifiers are inefficient at these frequencies. (For video modulation in television see the chapter on Television.)

The electrode supply voltages of the oscillator are suddenly increased from below plate-current cutoff to a constant value during the pulse to obtain constant oscillator frequency and power, and are then suddenly decreased to terminate the oscillation as rapidly as possible. Any modulator or driver tube is arranged to carry maximum current only during the pulse so as to conserve power, and particularly to allow it to

be pulsed to a peak current far beyond average current and power ratings. The maxima are limited by the safe electrode voltages for breakdown and by the average plate power.

Power oscillators are usually magnetrons for centimeter waves, or triodes for frequencies up to about 1,000 Mc. Magnetrons have a grounded anode or shell and the cathode is pulsed negatively for oscillation. Triodes are pulsed by applying a positive pulse to the grid or plate, by a negative pulse to the cathode, or by some combination of the above. A grid-pulsed triode is easily maintained in oscillation when once started and special means may be required to terminate the pulse.

Pulse-modulation systems include means for applying the requisite power to the oscillator tube and for limiting the pulse to a specified duration. The repetition rate is usually determined in separate equipment which supplies a short trigger pulse with a steep leading edge at the proper instant, although some systems run at a locally determined rate and supply triggering pulses for associated equipment.

The amplitude of the r-f envelope usually follows the pulse voltage closely except at the leading and trailing edges as exaggerated in Fig. 3. In triode oscillators, the

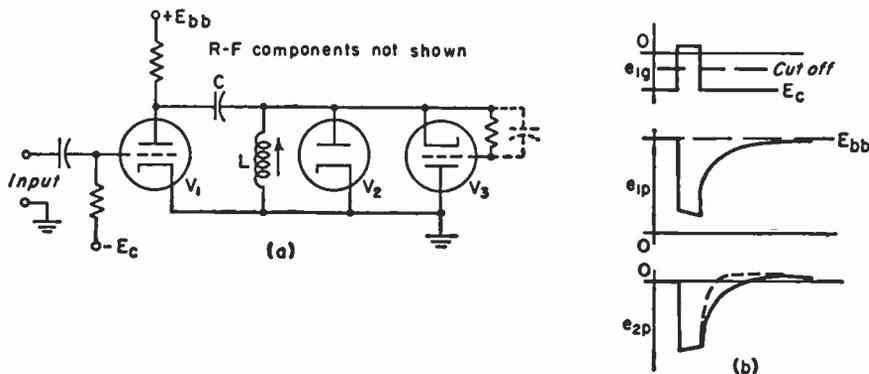


FIG. 26. Capacitance-coupled driver amplifier.

leading edge may be rounded if the Q of the r-f circuits is high and there is insufficient feedback to establish stable oscillating conditions quickly. The spike appears for excessive r-f feedback in triodes or insufficient modulator capability. It appears in magnetrons when the conditions for stable operation of the tank circuit are being established. There is a tendency for high- Q tank circuits to ring after the oscillator is cut off thereby showing a short logarithmically damped transient at the trailing edge of the pulse.

Only a few simplified circuits from many possible driver circuits are shown here. Comprehensive descriptions of the circuits used in radar are contained in "Principles of Radar."¹

30. Driver Amplifiers. A positive preformed pulse of the proper duration may be applied to the control grid of a triode or tetrode video power amplifier for driving a triode or magnetron oscillator. In Fig. 26a is shown a simplified circuit for coupling a driver V_1 to an oscillator V_2 through a capacitance C which is charged during the intervals between pulses. The plate of the driver tube (which may be several tubes in parallel) swings downward from E_{bb} during the pulse thereby providing adequate plate-cathode potential for the oscillator. The wave forms are indicated in Fig. 26b. Only a small part of the energy stored in C is discharged during the pulse. This energy is replaced through diode V_3 following the pulse, thereby returning the oscillator cathode to ground potential. The trailing edge of the modulating pulse may be more

¹ M.I.T. Staff, "Principles of Radar," McGraw-Hill, 1946.

quickly returned to zero by providing the *tail-sharpening* inductance L which builds up current in the direction shown by the arrow during the pulse. This current continues to flow into the circuit-to-ground capacitance after the pulse to restore normal conditions more quickly. The diode prevents the oscillator cathode from becoming positive with respect to ground.

The coupling from a driver to a magnetron oscillator may be through a special pulse transformer with a bifilar secondary winding (with or without a transmission

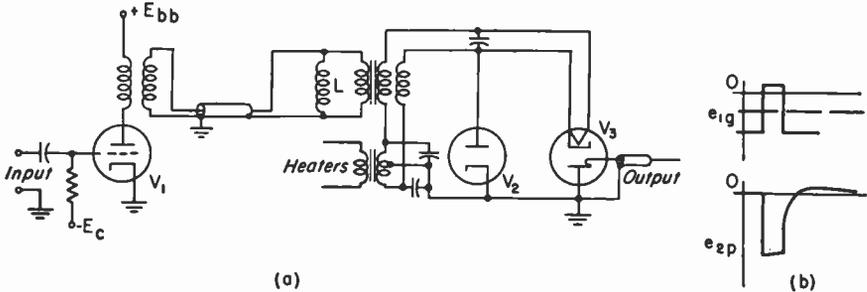


FIG. 27. Transformer-coupled driver amplifier.

line) as shown in Fig. 27a. A pulse transformer with a single secondary winding is used in a similar circuit for triode oscillators having grounded cathodes. The damping diode V_2 carries current during the first positive half cycle of transient oscillation following the pulse. A current builds up in inductance L (which alternatively may be on the secondary of the transformer) during the pulse and continues to flow immediately thereafter to charge the circuit capacitance quickly and restore normal conditions. Wave forms are shown in Fig. 27b.

Low-power triodes are sometimes pulsed by connecting the driver tube directly between oscillator cathode and ground as shown in Fig. 28. The cathode of the oscillator V_2 rises to E_{bb} between pulses while the supply potential for the grid is at the ground potential. A positive preformed pulse applied to the driver grid of V_1 causes the oscillator cathode to drop quickly to only a few volts above ground. Oscillation then starts quickly since the grid-cathode bias of V_2 is reduced as the plate-cathode potential is increased. The reverse operation quickly biases the oscillator beyond plate-current cutoff at the end of the pulse.

FIG. 28. Combined grid- and cathode-pulsed amplifier.

1. Clipping, differentiating, amplifying, and forming waves derived from sine waves of the desired repetition frequency
2. Triggering a one-shot multivibrator adjusted for pulses of the desired duration
3. Utilizing sharp wave fronts reflected in open-end or shorted-end pulse-forming networks having a definite delay
4. Triggering a blocking oscillator
5. Using negative feedback through a one-way delay line to terminate an otherwise longer pulse
6. Combining one or more of the above methods

31. Pulse Formation. The most common methods by which pulses are formed are:

The first two methods are suitable for pulses of 2 to 1,000 μsec or more duration. The other methods are used ordinarily for pulses of 0.1 to 10 μsec . The first method requires a large number of tubes for short pulses of reasonably rectangular shape and has been largely superseded.

32. One-shot Multivibrator. One of several possible types of one-shot multivibrator is the cathode-coupled circuit shown in Fig. 29 together with typical output wave forms.¹ It may be adjusted for pulses of 2 to 2,000 μsec duration by changing the capacitance C and/or the various resistors. Positive or negative pulses are available at the plates of V_2 and V_1 , respectively. A negative trigger may be used at the grid of V_2 instead of the positive trigger shown at the grid of V_1 . In any case the duration of the trigger should be less than the desired pulse, and no other extraneous pulses should be coupled to either grid.

Tube V_2 conducts between pulses since its grid-cathode voltage is approximately zero, while V_1 is cut off with a sufficiently high cathode voltage. A negative trigger

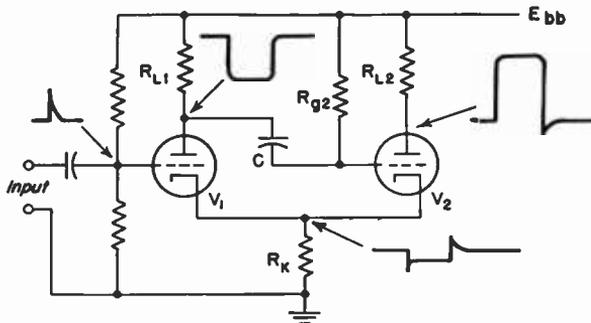


FIG. 29. Cathode-coupled one-shot multivibrator.

at the grid of V_2 reduces the current and the cathode voltage to a point where V_1 starts to draw current. Almost instantly the negative voltage developed in the plate load of V_1 drives V_2 to cutoff. The current in V_1 is then temporarily stable as determined by the tube characteristic, resistors R_{L1} and R_k , and the terminal voltages. The negative peak voltage at the grid of V_2 is substantially equal to the drop across R_{L1} although it is modified slightly by the inter-electrode and circuit capacitances. This grid voltage rises exponentially toward E_{bb} until V_2 again draws current. A positive voltage is regeneratively coupled almost instantly to the grid of V_2 via increased cathode voltage and reduced drop across R_{L1} so that V_2 momentarily draws more than normal current while V_1 is cut off. Grid current in V_2 then quickly reduces the grid voltage to normal, and the original conditions are restored.

The duration of the pulse depends on, but is less than, the time constant of C in series with R_{g2} and the combination of R_{L1} and the plate resistance of V_1 . The pulse may be lengthened by returning the grid resistor R_{g2} to cathode rather than to E_{bb} , since the voltage across the charging circuit for C during the pulse is then reduced. The duration of the pulse may be controlled to a limited extent by changing the size of R_{L1} since it determines the peak voltage to which C is charged during the pulse. Likewise this peak may be controlled by adjusting the tap of the voltage divider for the bias voltage applied to the grid of V_1 . The pulse may be lengthened by increasing C , R_{L1} , or R_{g2} , or by reducing the voltage at the grid of V_1 .

33. Pulse-forming Networks. Low-loss lines with distributed constants, open or shorted at the far end, or their network approximations, are used at various voltage levels for forming rectangular pulses of duration equal to the round-trip delay therein.

Figure 30a shows an open-end line with several sections of L and C approximating a uniform line. Figure 30b is a network simulating the sectionalized line more suited for operation at high voltage since, when completely charged, all the energy is stored in the single capacitance C_s , which is equal to the sum of the C s in Fig. 30a. The circuits L_1C_1 and L_2C_2 have natural frequencies equal to $1/t_w$ and $2/t_w$, respectively. C_s is also equal to $t_w/2R_c$, where R_c is the characteristic impedance of the line. A

¹ *Ibid.*, Chap. 2, Secs. 13-16.

pulse of half the initial voltage on the charged line appears across the load during the time the wave front travels to the open end and returns, when the line is discharged

through a resistance equal to the characteristic impedance R_c . The traveling voltage pulse is negative and is reflected from the open end of the line with the same polarity to cancel the remaining voltage at the load. The discharge is by smaller proportional steps when the load resistance is greater than R_c ; by decreasing alternately reversed steps when smaller.

Figure 31 shows a circuit in which a thyatron V_2 is made to act like a low-resistance electronic switch to discharge an open-end line by a positive trigger pulse on its grid. The network is desirably discharged through a load matched to the characteristic impedance. The by-pass diode V_3 provides current during any overshoot of voltage due to load mismatch so that the plate of V_2 can not become negative. The network recharges relatively slowly from the source E_{bb} through the charging diode V_1 and the inductance L after the pulse. This must be after the thyatron has deionized so that it will not short-circuit the source. The voltage on the line increases as the half cycle of an oscillatory charge to almost $2E_{bb}$, where it remains owing to the switching action of the diode V_1 until again triggered.

The charging diode and the inductance may be replaced by a high resistance at the expense of charging the line to only the voltage E_{bb} . A rotary-gap switch is sometimes used instead of the thyatron; also the charging may be from an a-c source.

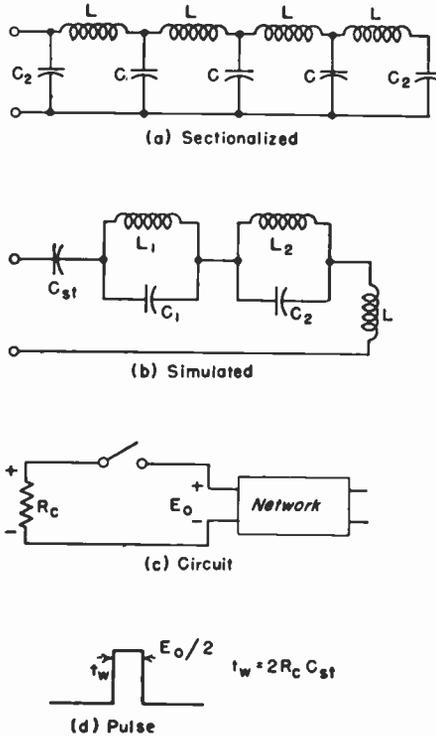


FIG. 30. Discharge of open-end delay network.

at the expense of charging the line to only the voltage E_{bb} . A rotary-gap switch is sometimes used instead of the thyatron; also the charging may be from an a-c source.

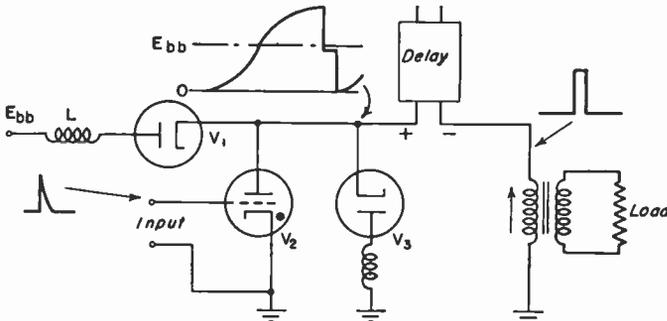


FIG. 31. Resonant charging of pulse-forming delay network.

Figure 32a shows a shorted-end line with sectionalized L and C . In this case t_w is $2L_{st}/R_c$. Figure 32b is a network which simulates the uniform line on short circuit.

When the line is connected to a source of constant current by opening a switch as in Fig. 32c, a pulse of voltage equal to $R_c I / 2$ appears until the initial wave travels to the end of the line and returns with reversed polarity to cancel the initial voltage, as shown in Fig. 32d.

Figure 33 shows a circuit in which such a shorted-end line may be used to develop a short pulse from a longer input pulse. When V_1 is suddenly made conducting by a wide input pulse, a negative voltage pulse appears at its plate for the duration of the round-trip delay of the line. The reverse occurs as the initial pulse ends. Tubes V_2 and V_3 may be used to clip the pulse by plate current cutoff so that only a rectangular pulse of the desired duration appears in the output. The approximate wave forms at each grid are shown. Tube V_3 acts as a cathode follower with low internal impedance and, therefore, has good regulation during the pulse.

34. Blocking Oscillator. An oscillator blocked after the first half cycle may be used to form a pulse. Either the first half cycle or the first positive overshoot of plate voltage due to transformer resonance may be used directly to modify the grid or plate supply voltage of an r-f oscillator. The pulse shape is then somewhat rounded, and better results are obtained after clipping in a driving tube. An iron-core pulse transformer couples the plate to the grid in proper polarity for oscillation which is prevented between pulses by a large negative grid voltage. The output load is

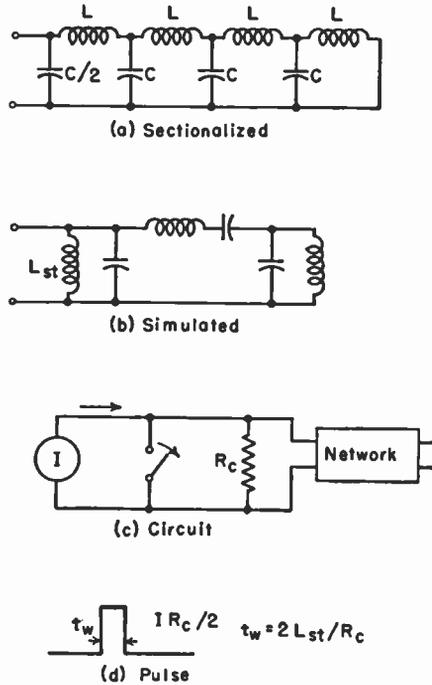


FIG. 32. Pulse form in shorted-end delay network.

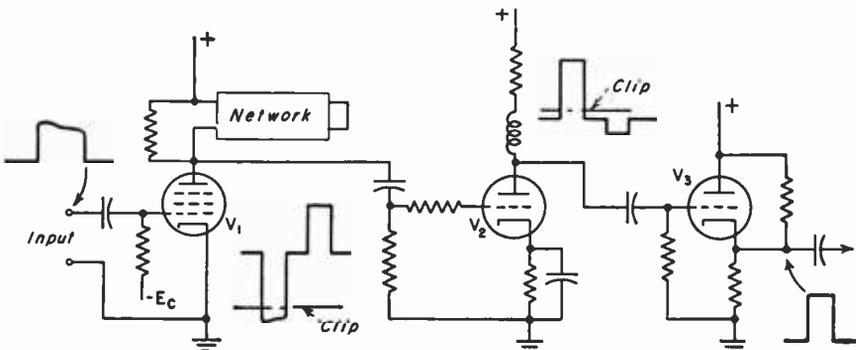


FIG. 33. Pulse forming by shorted-end delay network.

coupled from the same transformer. The plate current is abruptly increased by regenerative action when the grid voltage is made sufficiently positive, thereby

initiating a half cycle of violent oscillation. Bias voltage is produced thereby to prevent further regenerative operation until again triggered.

The voltage on C_2 in Fig. 34 is increased at each trigger pulse through C_1 and diode V_1 by an amount equal to $C_1/(C_1 + C_2)$ times the difference in voltage of the pulse and the voltage which is on C_2 immediately preceding the pulse. The tap for the cathode of blocking oscillator V_3 may be adjusted so that one or more input pulse steps are required to raise the grid potential above plate-current cutoff, thereby

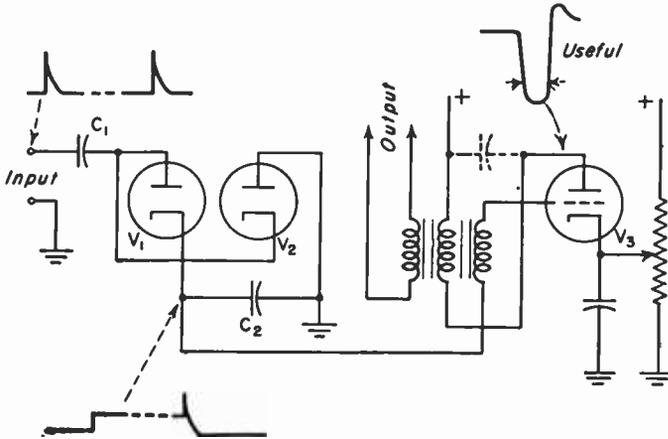


FIG. 34. Triggered blocking oscillator.

providing operation at the trigger repetition frequency, or some submultiple thereof. This expedient is useful where *division* of the repetition rate is desired, as, for example, when the average load in the equipment corresponding to a low repetition rate must not be exceeded. The first half cycle of a violently regenerative oscillation is started when plate current flows. Grid current is then drawn through diodes V_1 and V_2 in series, discharging C_2 so that the oscillation is not repeated until triggered again.

The grid current during each initial half cycle of the blocking oscillator of Fig. 35 charges C_1 to such a voltage that the tube then remains cut off for a considerable interval, depending on the time required to discharge C_1 through R_1 and R_2 in parallel. This determines the repetition rate and no trigger pulse is needed. Plate current is terminated at the end of each half cycle, and the first positive overshoot of plate voltage due to transformer resonance provides the useful output. This allows considerable storage of energy ($LI^2/2$) in the field of the transformer which develops a voltage

somewhat less than $I\sqrt{L/C}$ during the first damped half cycle after plate-current cutoff. The pulse duration is determined by the transformer inductance and circuit capacitance.

35. Termination of Pulse in Oscillator Circuit. The circuit for the supply of grid or plate voltage to triode r-f oscillators may be so arranged that the duration of the pulse, when once triggered, is determined in the oscillator circuit itself. Such arrangements are used at low power and where the pulse shape is not critical.

Figure 36 shows a circuit for a triode oscillator in which the pulse length is determined by a combination of the time constants of the grid circuit of the driving tube V_1 and of the cathode circuit of the oscillator V_2 . The cathode of V_2 is maintained at a sufficiently high potential between pulses by the bleeder to prevent the flow of plate current. The leading edge of the input pulse raises the oscillator grid voltage and initiates oscillation. C_2 then charges slowly to raise the cathode voltage while the grid supply voltage falls as the charge on C_1 leaks off through R_1 . After an interval

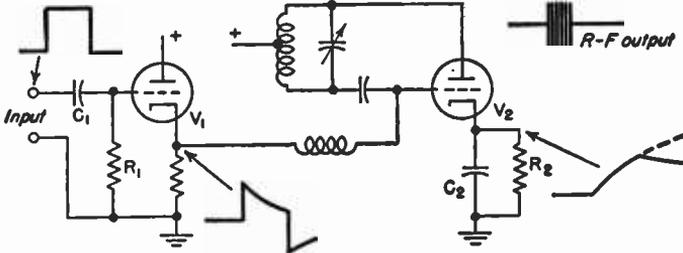


FIG. 36. Pulsed triode. Time constants of supply circuits determine duration.

which may be adjusted by changing C_1 , R_1 , or C_2 , the tube can no longer oscillate and the pulse is terminated.

Figure 37 shows a circuit for a triode oscillator in which the pulse length is determined by an open-end delay network in the cathode circuit. Tube V_2 is biased beyond plate-current cutoff between pulses and is started by a pulse slightly wider than desired, applied to the driver V_1 . The average oscillator cathode current initiates a pulse in the cathode delay network which is reflected from its open end with the same polarity and is returned to the cathode to raise its potential and terminate the pulse.

Alternatively, a delay network may be used in the grid circuit. In this case, a pulse of voltage negative relative to the grid, is initiated by the grid current at the start of

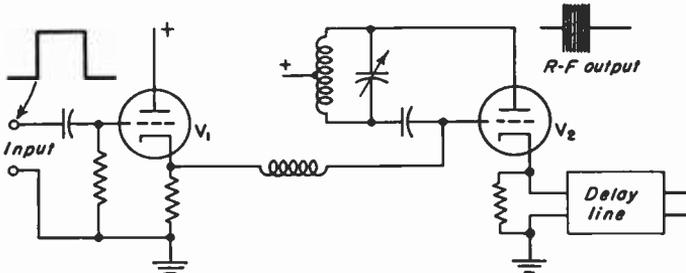


FIG. 37. Pulsed triode. Delay line determines duration.

oscillation. The pulse travels to the end of the line and returns with the same polarity to bias the grid beyond plate-current cutoff and terminate the pulse.

MODULATORS FOR P-M AND F-M WAVES

36. Phase Modulators. The usual method of producing p-m waves is to combine the output of a suppressed-carrier balanced amplitude modulator (see Sec. 24) with an unmodulated carrier which differs in phase by 90 deg from the original carrier. A vector diagram of the carrier and the net side-frequency components plotted relative to the carrier is shown in Fig. 38. The net side-frequency voltage E_s is in the direc-

filter whose output is proportional to the signal frequency before application to the control tube, p-m waves are produced.

38. Phasitron. A tube, called the *phasitron*, gives a large amount of phase modulation proportional to the current in a coaxial modulating coil surrounding it.¹ The tube utilizes a radial electron stream formed into a disk with a wavelike edge. The scalloped beam is made to progress at a constant average angular velocity about the cathode by deflecting electrodes excited by a crystal oscillator, and alternately to strike anode areas connected to the opposite terminals of a balanced output circuit. The phase of the scalloped edge is advanced or retarded by the modulating field.

When modulating voltages of constant amplitude are applied to the coil, the current is inversely proportional to their frequency and f.m. is obtained directly. The phase shift over which the modulation is linear may be more than 360 deg so that less frequency multiplication is required to obtain the standard system deviation for f-m broadcasting than with conventional phase modulators.

FREQUENCY CONVERTERS AND MIXERS

39. Class A Frequency Converters. A power-series expression for plate current accurately determines the output of a tube used in double-modulation service when the plate current is not swung to cutoff.

The signal and local oscillator voltages e_s and e_o may be applied to the same or separate grids. The third term of the series A_2e^2 of Eq. (48) when expanded yields the i-f plate current,

$$i_i = \frac{A_2 E_0 E_c r_p (1 + m \cos at) \cos (\omega_0 \pm \omega_c)t}{r_p + Z} \quad (72)$$

when the signal and oscillator voltages are

$$e_s = E_c (1 + m \cos at) \cos \omega_c t$$

and

$$e_o = E_0 \cos \omega_0 t$$

and r_p and Z are the plate and output impedances, respectively.

The impedance of the output circuit at the i-f frequency is usually too large in comparison with the plate resistance to be neglected. It is desirable to use a low-impedance primary winding in the output circuit of triodes to obtain the best conversion gain.

The output of a class A frequency converter is low since the capability of the tube is not realized fully. It is a van der Bijl modulator with a change in frequency between the input and output carriers. When terms of higher order than A_2e^2 are negligible, the operation is according to a square-law characteristic, and there are no spurious intermodulation responses. The output voltage may then be expressed as

$$e_i = \frac{Z r_p A_2 E_0 E_c}{r_p + Z} (1 + M) \cos \omega_i t \quad (73)$$

$$= \frac{Z r_p}{r_p + Z} S_c E_c (1 + M) \cos \omega_i t \quad (74)$$

where M represents the instantaneous modulation. S_c is the *conversion transconductance* which is the i-f plate current per volt of applied signal for the selected oscillator and bias adjustments.

Conversion gain is the ratio of the voltage developed in the i-f circuit (usually measured at the grid of the following tube) to the signal voltage applied to the converter tube.

40. Superheterodyne Frequency Converters. Frequency converters, sometimes called *first detectors*, in superheterodyne receivers are operated with as large a local

¹ Adler, R., A New System of Frequency Modulation, *Proc. IRE*, January, 1947, pp. 25-31.

oscillator voltage as possible without endangering nonlinear i-f response. This ensures the highest conversion transconductance and conversion gain.

When the signal and oscillator voltages are applied to the same grid, the tube is biased nearly to cutoff in the absence of oscillator voltage, and the oscillator voltage is made a volt or so less than that which would cause grid current. Plate current then flows for the positive peaks of oscillator voltage and is cut off for a large part of the cycle. The plate current is modulated by a relatively small signal voltage, and the sum or difference frequency components (usually the latter) are selected and tuned in the plate circuit as the i-f output.

The conversion transconductance under the most favorable conditions does not exceed about 0.3 of the transconductance of the same tube as an amplifier. Limited a-v-c bias may be applied to the signal grid for control of the conversion gain of variable- μ tubes.

41. Special Converter and Mixer Tubes. Interaction between the signal and oscillator circuits of frequency converters produces undesirable oscillator detuning. This may be reduced somewhat by coupling the oscillator voltage to the suppressor grid or to the cathode of a pentode converter tube, but even these expedients are

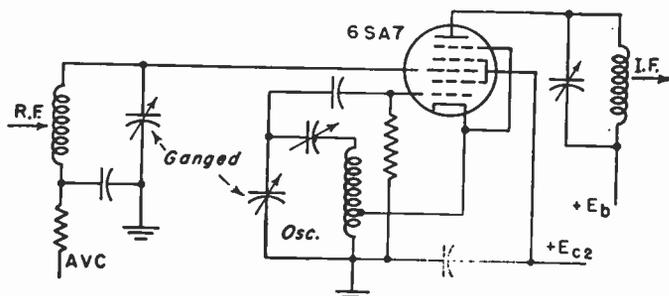


Fig. 40. Typical pentagrid-converter circuit.

ineffective when the percentage difference between the signal and oscillator frequencies is small, as in the h-f bands of so-called *all-wave* broadcast receivers.

Special pentagrid mixer tubes, such as the 6L7 tube, have been designed for frequency-converter service which give superior performance due to better shielding between the signal and oscillator grids, high plate resistance, high conversion transconductance, and suitability for a.v.c.

Specially designed multigrad converter tubes of several types are also available in which two of the electrodes serve as the grid and plate of the oscillator circuit. The electron stream as initially controlled by the oscillator is modulated by the signal applied to a screened control grid.

A typical pentagrid-converter circuit for a standard or short-wave broadcast receiver is shown in Fig. 40.

The best stability and conversion gain for u-h-f signals are obtained with a high i.f., with a separate oscillator tube, and with the signal and oscillator voltages applied to the same grid of the converter tube.

42. Spurious Responses. The shot noise in a converter tube contributes considerably to the noise figure of a radio receiver. The noise from a particular tube as a converter is always three or four times as great as when the same tube is used as an amplifier. The noise is sometimes expressed in terms of an equivalent grid resistance whose thermal noise would account for the observed noise in the output. The equivalent resistance (and noise) is less in triodes and in high-gain pentodes than in the special converter tubes whose main advantage is the incorporation of the electrodes for both oscillator and converter in one envelope.

The desired signal input applied to converter tubes must not be sufficient to draw signal grid current or to drive the plate current to complete cutoff for the positive peaks of oscillator voltage during maximum modulation. Too large a signal produces harmonic distortion of the envelope of the i-f carrier.

The linearity may be checked by measuring the i-f output for a range of unmodulated signal voltages for the oscillator and bias conditions selected. For full modulation the linearity is satisfactory to half the signal level indicated by the test. Such a test is analogous to that described in Sec. 21 for modulator tubes, but in this case it is difficult to develop a sufficiently high plate-circuit impedance at a low test frequency to simulate the value at i.f.

Strong interfering signals of a number of definite frequencies produce spurious responses in the output. A signal at the converter input, either higher or lower in frequency than the oscillator by the i.f., produces equal response in the tuned output circuit. One of these having been selected as the desired signal, the other is known as the *image response* and must be attenuated in preselector circuits ahead of the converter tube.

Other spurious responses are due to the terms A_3e^3 , A_4e^4 , etc., in the expression for plate current. These terms are absent in a tube with a square-law characteristic but are present in a normally operated converter tube for which the series converges slowly.

With insufficient preselection at high levels of interfering signal, an interfering program response may be heard, e.g., when the second harmonic of the interfering signal carrier frequency differs from the fundamental of the oscillator frequency by the i.f. When a desired signal is also present, a beat between the intermediate frequencies from the two signals is rectified in the second detector.

This spurious response is predicted from the expansion of the term A_3e^3 where

$$e = E_s \cos \omega_s t + E_i \cos \omega_i t + E_o \cos \omega_o t \tag{75}$$

where E_s and E_i are the amplitude coefficients at the signal and interfering frequencies. The desired output is

$$A_2 E_s E_o \cos (\omega_o - \omega_s) t \tag{76}$$

and the particular interfering output is

$$\frac{3}{2} A_3 E_i^2 E_o \cos (2\omega_i - \omega_o) t \tag{77}$$

Other interferences may be predicted from the other terms of the expansion of A_3e^3 , A_4e^4 , etc.

43. Crystal Converters. Silicon crystals in small cartridges of standard size are used extensively as converters for centimeter waves. A

typical placement is shown diagrammatically in Fig. 41a where the crystal is at the end of a resonant concentric line to which the signal and oscillator voltages are applied. The base of the crystal is insulated by a low-inductance by-pass capacitor. The output circuit is tuned at the i.f. as shown by the equivalent circuit of Fig. 41b. It is designed to conform to the low effective impedance of the crystal which operates from the equivalent of a series-tuned r-f circuit.

Since a crystal is not an amplifier there can be no conversion gain. The *conversion*

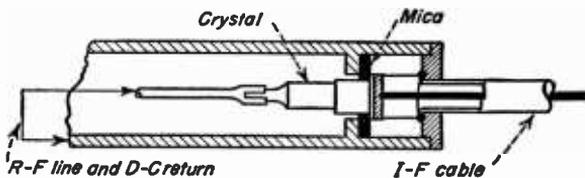


FIG. 41a. Crystal mounting.

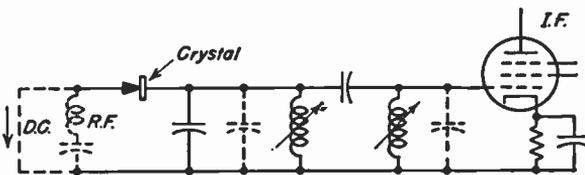


FIG. 41b. Equivalent crystal circuit.

loss is taken as the ratio of the available i-f signal power to the available r-f signal power. It varies with the circuit impedance, but is usually about 6 to 10 db.

The *noise-power ratio*, usually from two to four times, is taken as the ratio of the available noise power in the i-f circuit to the available noise power when a specified resistance is substituted.

Crystals are easily damaged, and voltages should not be applied which are greater than about 5 volts in the blocking direction or which result in more than about 1 volt d.c. in a resistive load.

A-M DETECTORS

44. Diode Peak Detectors. A diode so connected to a carrier-frequency source that it is effectively in series with a capacitor to provide a low-impedance path at that frequency, while the path for the rectified current is through a load resistor which

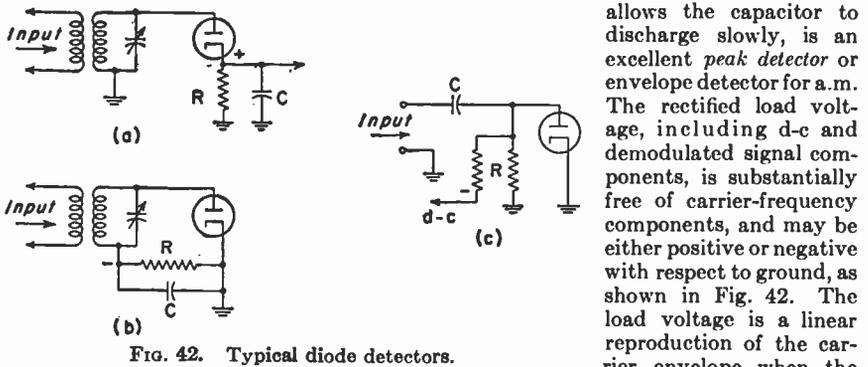


FIG. 42. Typical diode detectors.

allows the capacitor to discharge slowly, is an excellent *peak detector* or *envelope detector* for a.m. The rectified load voltage, including d-c and demodulated signal components, is substantially free of carrier-frequency components, and may be either positive or negative with respect to ground, as shown in Fig. 42. The load voltage is a linear reproduction of the carrier envelope when the carrier voltage is sufficiently high. This is of the order of 0.5 volt or more for usual tubes. Diode current flows only for an instant at the peak of the carrier voltage in the forward direction, when the net voltage across the diode is positive and the internal resistance is small compared to the load resistance. The pulses of current charge the series capacitor to nearly the same voltage as the carrier envelope and bias the diode beyond cutoff during the remainder of the cycle when the internal resistance is very high.

The charge, which is replenished at each pulse, must leak off sufficiently before the following pulse for the bias voltage to follow the carrier envelope at its maximum slope. The critical relation is

$$\frac{1}{aRC} \geq \frac{m}{\sqrt{1 - m^2}} \tag{78}$$

where $a/2\pi$ = modulation frequency

R = load resistance

C = by-pass capacitance

m = degree of modulation

C is made large enough to by-pass the carrier but small enough to reproduce the modulation. The load resistor should not be shunted by impedances which change its effective value in the frequency range of the demodulated signals, as discussed in Sec. 45.

Figure 43 shows the rectification characteristic of a typical 6H6 diode. Rectified current is given for several steps of rms input voltage in terms of d-c load voltage. These characteristics may be determined by test at 60 cycles according to the method of Sec. 21.

The ratio of the d-c voltage E_d to the peak value of the applied voltage E_a is the voltage rectification efficiency of the diode.

$$\eta = \frac{E_d}{E_a} \tag{79}$$

Curves for constant efficiency plotted on Fig. 43 would have the same general characteristic slope as the load-resistance lines shown.

Figure 44 shows a load rectification diagram for a 100,000-ohm load resistance as taken from Fig. 43 or plotted directly from test data. The linearity of Fig. 44 indicates that a diode is suitable for signals as large as can be supplied by the previous amplifier without overloading. The slight departure from linearity at the origin shows a small amount of inherent distortion for small input signals.

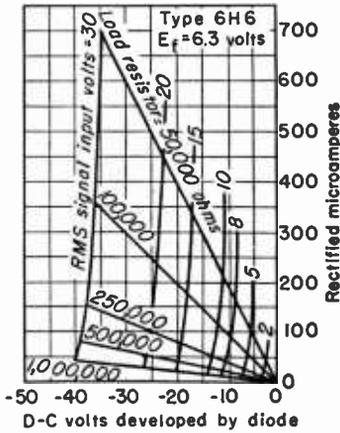


FIG. 43. Rectification characteristics of diode.

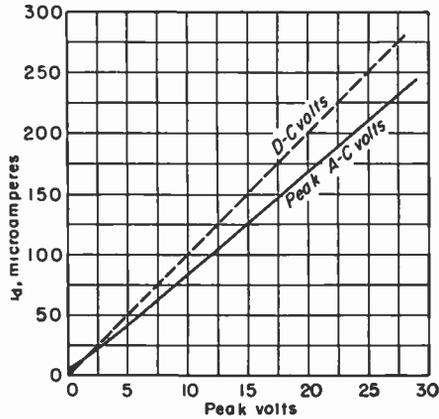


FIG. 44. Load rectification diagram of diode.

The difference between the peak a-c and the d-c voltages is almost proportional to the rectified current I_d and, except with small input signals, is equivalent to a resistance drop in series with the output. This effective internal diode resistance R_d is

$$R_d = \frac{E_a - E_d}{I_d} \tag{80}$$

A Fourier series expansion of the pulses of current shows that, when the load resistance is large, the peak value of the fundamental component at carrier frequency is twice the d.c. in the load with less than 1 per cent error. The effective input resistance of the diode is then

$$R_i = \frac{E_a}{2I_d} = \frac{E_d}{2\eta I_d} = \frac{R}{2\eta} \tag{81}$$

The impedance of the source Z_0 produces a drop in voltage

$$Z_0 I_a = 2Z_0 I_d \tag{82}$$

and the equivalent generator peak voltage E_0 is

$$E_0 = I_d(R + R_d + 2Z_0) \tag{83}$$

The above equations apply to an unmodulated carrier. In a practical circuit the d-c load may necessarily be shunted by other impedances, much higher than the load resistance at the modulation frequency. Figure 45 shows a circuit in which R_s is a decoupling resistor for a-v-c supply and R_1 is an output resistor blocked for d.c. The

resistor R_1 , with the by-pass capacitors C_1 and C_2 , filters the h-f carrier components from the output. The d-c load resistance R is

$$R = R_1 + R_2 \tag{84}$$

and the a-c load impedance R' at low modulation frequencies is

$$R' = R_1 + \frac{R_2 R_3 R_4}{R_2 R_3 + R_2 R_4 + R_3 R_4} \tag{85}$$

The ratio R'/R is called the a-c/d-c ratio and should be as near unity as possible.

The modulation-frequency voltage E_m developed in the load by the modulation-frequency component of the load current I_m is

$$E_m = R' I_m \tag{86}$$

The corresponding generator side-band voltage E_s relative to the carrier voltage is

$$E_s = I_m (R' + R_d + 2Z_o') \tag{87}$$

where Z_o' is the impedance of the source at the side-band frequency.

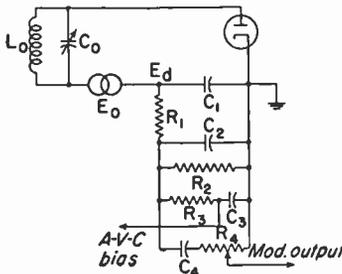


FIG. 45. Typical diode circuit.

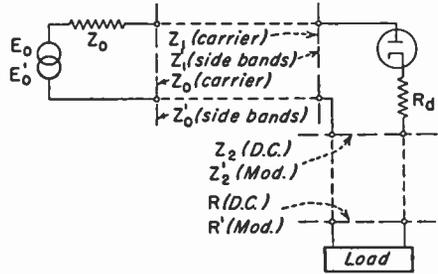


FIG. 46. Equivalent diode circuit.

The equivalent output impedance Z_2 of the diode is

$$Z_2 = 2Z_o\eta \tag{88}$$

for the d-c component and

$$Z_2' = 2Z_o'\eta \tag{89}$$

for the modulation frequencies.

The diode acts like a motor generator with input and output impedances which depend on the connected output and source impedances. An equivalent diagram is shown in Fig. 46.

45. Diode Performance. The approximate relations between the rectified current and the peak voltages with resistive load during modulation are shown in Fig. 47. The impedance of the source to the side-band frequencies Z_o' is here assumed to be the same as for the carrier Z_o . The ordinates of the upper part of the diagram are for d-c or modulating-frequency load current. The abscissas are for d-c or a-c peak voltages. The load voltages are the current multiplied by R for d-c values and by R' for modulating-frequency values. The drop in the diode and in the source is the current multiplied by $(2Z_o + R_d)$. The envelopes of the input, source, and voltage are shown for one cycle of full modulation and for one cycle of modulation which just reaches diode cutoff during inward modulation. Specific dimensions are indicated for one point in the envelope in Fig. 47. The figure is exaggerated for purposes of illustration. The output voltage across resistor R_4 in Fig. 45 is reduced further by the voltage drop in resistor R_1 (not shown).

The peak of output voltage is clipped and the envelope of the input voltage rises when the degree of instantaneous applied inward modulation of the source exceeds the value

load resistances. Lack of balance is not so serious at the low-impedance levels used in detection for television.

48. Grid Detectors.

This form of detector was used mainly before diode detectors were known and is, therefore, largely only of academic interest today. The circuit elements connected between the grid and the cathode of a grid detector act substantially the same and are determined by the same considerations as are the corresponding elements in a diode. As with a diode the load on the previous circuit is equivalent to a shunt load of approximately half the resistance of the grid leak.

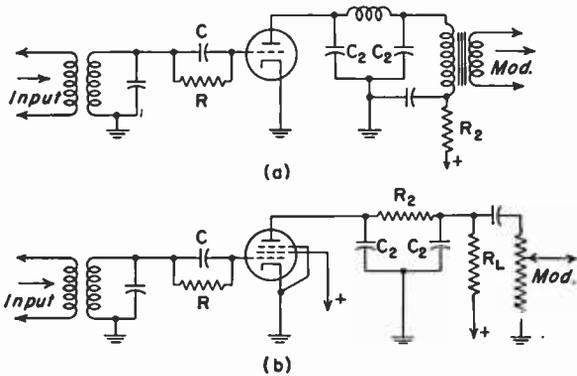


FIG. 49. Power-grid circuit detectors.

Figures 49a and b show typical triode and pentode power-grid detector circuits.

The tube must be operated with low plate (or screen) voltage since in the absence of signal there is no bias voltage. This limits the useful plate swing. The range may be extended by operating the tube from a high-voltage *B* supply with resistance coupling, or with a resistor in series with the load, by-passed for audio, with transformer coupling.

The carrier-frequency components are filtered from the load circuit by *LC* or *RC* networks.

The operation is illustrated in Fig. 50 which shows the dynamic characteristic including the plate load selected. This may be obtained at any convenient frequency by the method of Sec. 21.

A pulse of grid current charges the grid capacitor at each positive peak of grid voltage and establishes a negative grid bias

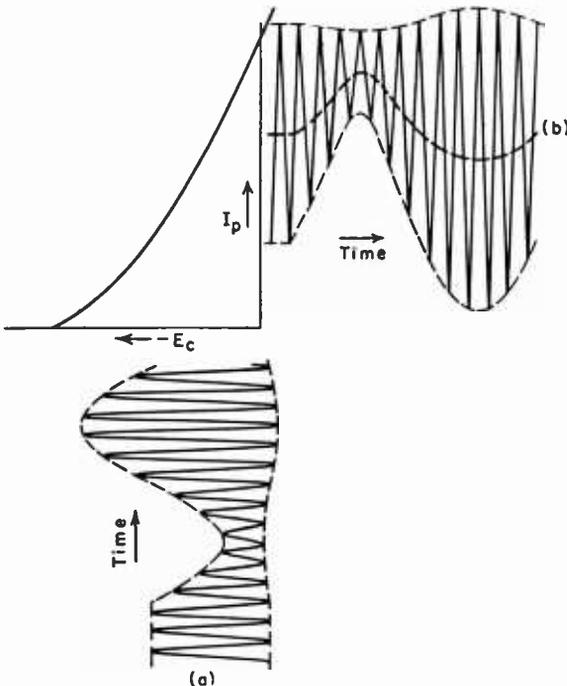


FIG. 50. Power-grid detection.

with the same losses relative to the envelope as in a diode. The net applied-grid voltage, after the loss (not shown) in the source impedance is deducted, is indicated by

curve *a* in Fig. 50. The instantaneous grid voltage due to the modulated wave is superimposed on the bias and produces instantaneous values of plate current as shown in curve *b* of Fig. 50. The amplified useful output is proportional to the average of the plate current upon which the individual cycles are shown superimposed.

If the negative peaks of instantaneous grid voltage swing over the curved lower portion of the characteristic, the audio is reduced and distorted by partial plate detection. This effect is exaggerated in Fig. 50, which shows more than normal curvature of the grid-plate characteristic to illustrate even-harmonic distortion of the modulation frequency in the output.

The audio output voltage which may be obtained satisfactorily is about 0.3 to 0.4 of the corresponding value for the same tube when used as an amplifier.

Power-grid detectors operate over a limited range of voltages which is insufficient for a.v.c.

49. Square-law Detectors. Square-law detectors were used commonly in broadcast receivers at an early period before tubes and circuits were available for developing sufficient carrier voltage for high-level grid or plate detectors. The sensitivity of a grid-leak detector is greatest for very weak signals when provided with a high-resistance grid leak. The grid then operates on a portion of the grid-cathode characteristic which is substantially square law and over which the grid current is never cut off. This method introduces harmonic distortion with high degrees of modulation and produces an output proportional to the square of the carrier voltage.

When the applied voltage is

$$e = E_c(1 + m \cos at) \cos \omega t \tag{98}$$

the power-series expression for grid current i_g yields from the term A_2e^2 for square-law detection

$$i_g = \frac{A_2E_c^2}{2} \left(1 + 2m \cos at + \frac{m^2}{2} + \frac{m^2}{2} \cos 2at \right) \tag{99}$$

The audio component i_a of the grid current is

$$i_a = A_2E_c^2 \left(m \cos at + \frac{m^2}{4} \cos 2at \right) \tag{100}$$

The same audio components appear in the output. For full modulation the second harmonic distortion is then 25 per cent.

A tube biased to a curved portion of its grid-plate characteristic operates approximately as a square-law detector for low input voltages. The output and distortion vary in the same way as with the grid-leak detector for weak signals.

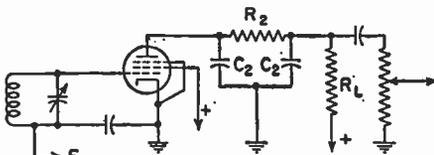


FIG. 51. Circuit for plate detection.

However, the weak-signal plate detector does not load the previous circuit and is therefore suitable for use as a vacuum-tube voltmeter.

50. High-level Plate Detectors. A plate-circuit detector for large signals is biased nearly to cutoff in the absence of a signal, and a high plate voltage is used to extend the range of operation. A typical circuit is shown in Fig. 51.

The operation is illustrated in Fig. 52 in which plate current is plotted against bias voltage for the load impedance selected. The positive excursions of the instantaneous grid voltage produce substantially half waves of plate current, the average values of each pulse producing the audio voltage in the load, while the h-f components are by-passed to ground. Curve *a* of Fig. 52 shows the instantaneous values of input signal applied to the grid which is biased by the voltage $-E_c'$. Curve *b* shows the instantaneous values of plate current superimposed on the demodulated output.

The effect of the curvature of the grid-plate characteristic is exaggerated in this figure to illustrate the even harmonic distortion from this cause in the output.

The power-series expression for plate current converges too slowly for analytical purposes, and the performance is determined by test. The linearity of the output vs. the input is shown in Fig. 53 which is a load-rectification diagram for a pentode plate detector. This diagram may be obtained by test at 60 cycles when the impedances are made the same as in actual use. Load current is plotted against rms values of input voltage for the selected load and bias conditions.

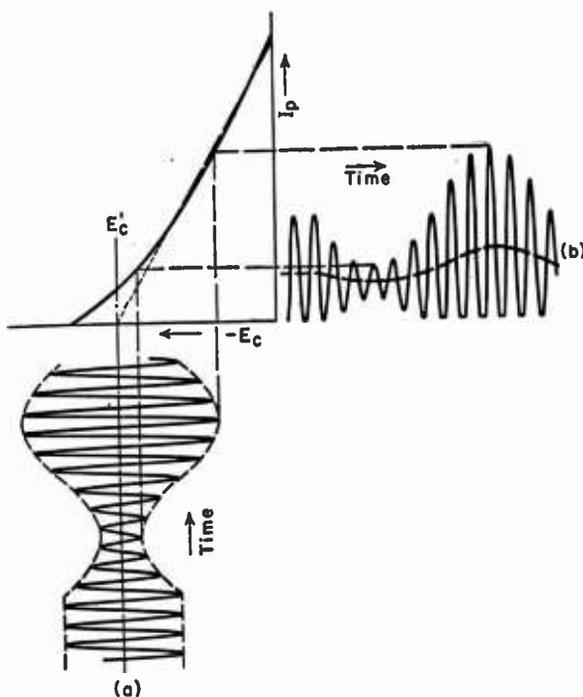


FIG. 52. Analysis of plate-circuit detection.

The intercept of the extension of the linear portion of the grid-plate characteristic with the E_c axis indicates the approximate value of bias voltage for maximum output and minimum distortion, as shown by the dotted line in Fig. 52.

When the tube is driven hard enough to draw grid current, the source impedance must be low or grid rectification will reduce and distort the output. Pentodes may be biased to draw no grid current over the working range.

The degree of modulation which may be handled without distortion is limited and may be calculated from the linear portion of the selected curve of Fig. 53. A slight improvement in the performance over a range of input voltages may be obtained by increasing the bias for large signals.

While the performance of a plate detector is indicated by the curves of Fig. 53, the following detector quantities are often used. The detection plate resistance R_d is

$$R_d = \left. \frac{\partial E_p}{\partial I_p} \right]_{E = E_c} \tag{101}$$

evaluated under operating conditions with a carrier E_c (or its equivalent at 60 cycles) applied to the grid. It replaces R_p in equations for tube circuits involving detectors. The conversion transconductance S_c is

$$S_c = \left. \frac{\partial I_p}{\partial E} \right]_{E=E_c} \tag{102}$$

The efficiency of detection D is

$$D = \frac{S_c R_d}{\mu} \tag{103}$$

where μ is the amplification factor and is also evaluated under operating conditions.

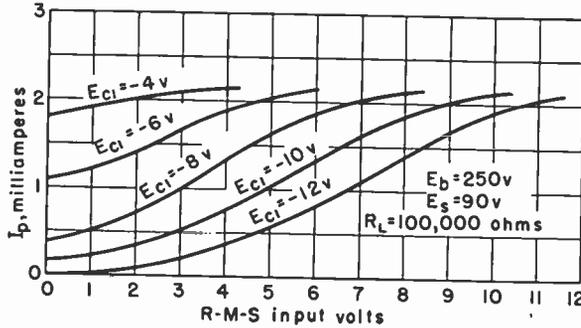


FIG. 53. Load-rectification diagram of pentode.

The change in plate current due to modulation is

$$\Delta I_p = \frac{\mu D m E_c}{Z_L + R_d} = \frac{S_c m E_c R_d}{Z_L + R_d} \tag{104}$$

where m = degree of modulation
 Z = load impedance

51. Infinite-impedance Detector. A triode self-biased nearly to plate-current cutoff by a large cathodic resistance (by-passed for the carrier frequency) passes pulses of plate current at the positive peaks of grid voltage. These pulses act in the RC circuit of a diode with the exception that the energy is obtained from the plate circuit.

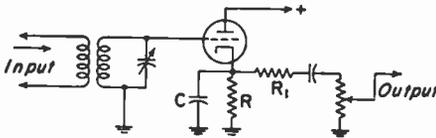


FIG. 54. Infinite-impedance detector.

The grid does not draw current and, therefore, does not load the preceding circuit, so that the over-all voltage efficiency may be about 0.8 to 0.9, or higher than in a diode.

The bias increases with carrier voltage and follows the modulation up to the limit where the degree of modulation is

$$m = \frac{I_0 + I_s}{I_s \sqrt{R^2 a^2 C^2 + 1}} \tag{105}$$

where I_0 = plate current for zero signal
 I_s = increase in plate current with signal

A proper choice of R and C , therefore, permits full modulation without peak clipping.

The shunting effect of any impedance in the grid circuit of the following tube can be made negligible, since R is smaller than in diode circuits.

The operation is linear for high levels of carrier input up to an output level limited only by the plate-supply voltage. A disadvantage of this circuit is its inability to supply voltage for conventional a-v-c circuits. This form of detector was not proposed until a rather late date when high-level amplification could be attained readily and has not been used widely in broadcast receivers.

52. Regenerative and Heterodyne Detectors. A detector with controlled positive r-f feedback is regenerative and is capable of very great amplification of the signal before detection. The tube also acts as a conventional plate detector by operating on a curved portion of its characteristic. The regeneration may be controlled by adjusting the reactive coupling in the feed-back circuit or by altering the transconductance of the tube. Examples of control are the adjustment of the screen voltage in pentodes, and the adjustment of the coupling of the feedback winding in triodes, as shown in Fig. 55. Oscillation is produced when the gain through the tube and feed-back loop is greater than unity in the phase to reinforce the input voltage.

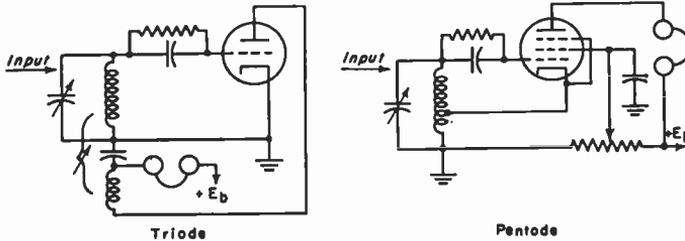


FIG. 55. Regenerative and heterodyne detectors.

The regenerative gain greatly increases the r-f tuned circuit voltage when the adjustments are for as much feedback as possible without oscillation, particularly at low input levels. The adjustments are critical near the spill-over point since the gain is inversely proportional to the $1 - \mu\beta$ of the usual formula for feedback.

The same circuit may be used in the *heterodyne* or *beat-frequency* reception of c-w telegraph signals when adjusted to oscillate weakly, although the frequency of the audio beat note is critically dependent on minor variations, such as in the supply voltage. The oscillator is not quite synchronized by the signal when operating properly. Undesired a-f *threshold howl* which may spoil the more sensitive adjustments may be avoided in triodes by resistance coupling to the audio tube or by shunting any inductive plate load with a low resistance. Threshold howl is not troublesome in pentodes since the plate current is substantially independent of the load impedance.

The envelope of a c-w signal beating with a local oscillator is as expressed by Eq. (18). A beat note consisting of a fundamental plus percentages of its higher harmonics decreasing with their order is obtained by rectification by a separate linear detector. Only the fundamental is present in the output of a square-law detector such as is provided approximately by the plate circuit of the oscillator tube itself.

53. Superregeneration. A superregenerative circuit is usually connected as a triode which would oscillate continually if provided with sufficiently high electrode voltages. The oscillation is made intermittent by controlling the grid or plate voltage at a *supermodulation* or *quench-frequency* rate. The latter is always made at least twice the highest modulation frequency, and it may be several hundred kilocycles when used in u-h-f reception. The demodulated output appears in the plate circuit of the tube which inherently operates as a plate-circuit detector, or a separate linear detector may be used for the r-f voltage that appears across the tuned circuit.

The tuned circuit is effectively shunted by a conductance which is made alternately positive and negative by the quench mechanism. During the positive-quench half cycles (conductance negative) the circuit is regenerative and the r-f voltage builds up

from a low level. It decays during each negative-quench half cycle (conductance positive). The circuit may be made strongly regenerative by the fixed circuit potentials in the absence of quench voltage, thereby reducing the net circuit conductance to substantially zero. The added quench voltage then may be considered to vary the conductance almost equally positive and negative. When analyzed in this manner, the net over-all gain is the simple regenerative gain for the assumed initial condition plus the superregenerative gain owing to intermittent growth and decay of the circuit oscillations.

The voltage builds up from the thermal noise level of the tuned circuit in the absence of a signal. This random noise is then amplified and detected as a loud characteristic hiss. A signal coupled to the tuned circuit raises the initial level and the build-up is earlier in the quench cycle. The circuit is most sensitive to the effect of a signal during the period when the circuit oscillations are at their lowest level since, after build-up starts, the signal loses control until the next sensitive period. Pulses of amplitude and sensitivity vs. time alternate.

It is necessary that the positive conductance and the duration of the decay period be sufficient for the envelope to decay to the initial noise level in each quench cycle, or the accumulative effect from cycle to cycle will cause the level to depend on an integrated rather than an instantaneous response to the signal. The true signal must be sampled once each quench cycle. Any coupling to other circuits must be so damped

that the decay is not delayed by reflections therefrom. For example, an improperly terminated antenna transmission line may reflect waves which lengthen the decay time in an otherwise satisfactory unit.

The envelope of the r-f voltage across the tuned circuit consists of signal-modulated pulses repeated at the quench-frequency rate. Consequently, the demodulated output, filtered only for the r-f components, contains terms at each of the integral multiples of the quench frequency, including zero, which are modulated in amplitude by the signal. Any one of these components may be further amplified for final utilization. An audio transformer in the plate circuit of the tube may be used for coupling to an output amplifier when the

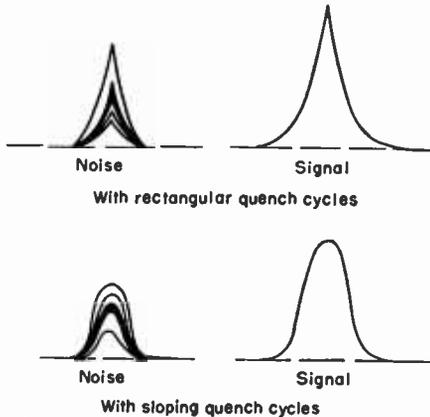


FIG. 56. Rectified pulses at quench frequency. Linear mode.

modulation is at a.f., since it carries currents at the signal frequency.

A superregenerative circuit may operate in linear or logarithmic mode. In the linear mode, the period of build-up is made so short that equilibrium of the circuit as an oscillator is not reached. The peak amplitude of each pulse is then directly proportional to the sampled signal. The rate and total amount of the build-up depend critically on the rate of change of circuit conductance from positive to negative. The build-up is logarithmic for rectangular quench cycles but takes a rounded form when the change of conductance is gradual. Rectified envelopes for noise and signal, with only r-f filtering, are illustrated in Fig. 56. The r-f voltage may change by a ratio of 1,000,000:1 during the quench cycle so only a small portion of the envelope is distinguishable in the linear plot of Fig. 56. The noise peaks in the output are of random amplitude since the random thermal noise is sampled at definite intervals. The multiple lines in the figure show this while the maximum density of lines shows the output at the rms noise level.

In the logarithmic mode, the positive quench half cycles are made long enough for

the oscillation to reach equilibrium after each build-up. The negative quench half cycle need only be long enough for the envelope to decay below the noise level. The rectified cycles, with only r-f filtering, are illustrated in Fig. 57. The output signal is due solely to the variation in the leading edge of the pulse and is proportional to the logarithm of the applied r-f signal envelope. This mode of operation provides an a-v-c action and limits the response to short external noise pulses of large amplitude. Downward a.m. is accentuated while outward a.m. is diminished, giving considerable distortion on high percentages of modulation.

The quench-frequency oscillator may be separate from the superregenerative tube, or the latter may be made *self-quenching* in a linear mode by allowing it to block at a superaudible rate. The size of the grid resistor is increased for this purpose. The build-up time is shortened by the signal input in self-quenched circuits, while the decay time remains constant. The result is that the quench frequency increases with the strength of the applied signal wave. The useful output is proportional to the logarithm of the applied r-f signal envelope.



FIG. 57. Rectified pulses at quench frequency. Logarithmic mode.

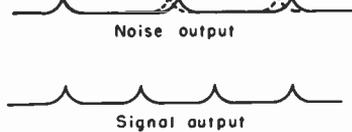
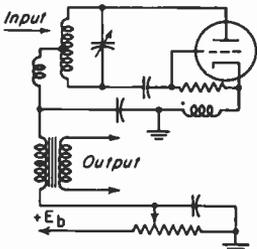
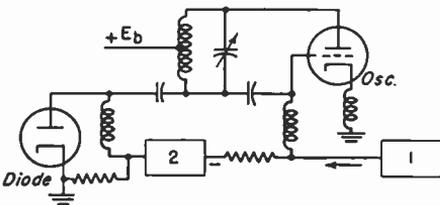


FIG. 58. Self-quenched superregenerative circuit.

Superregenerative circuits are of most value at frequencies where the side frequencies of the desired modulation cover only a small part of the acceptance band of the circuit. The effective band width and selectivity of the circuit may be obtained by the Fourier integral from the amplitude pulse. The band width is greatest for the sharp pulse obtained with a rectangular quench cycle. It is reduced by broadening the pulse as with gradual transition of circuit conductance. It increases with the quench frequency when a sine-wave quench cycle is used. It is substantially independent of the extremes to which the conductance is varied except for rectangular quench. It is inversely proportional to the capacitance of the tuned circuit. In practical cases it may be from five to ten times the quench frequency.

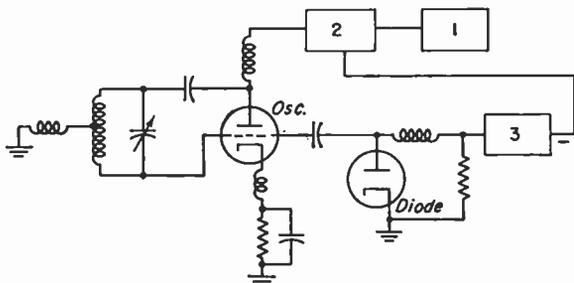


- 1 Quench-frequency oscillator
- 2 Quench-frequency amplifier and rectifier for a-g-s control

FIG. 59. Grid-quenched superregenerative circuit with automatic gain stabilization.

The audible noise in a superregenerative circuit is greater than in the simple circuit by the ratio of the r-f band width to the quench frequency since the noise components in the over-all pass band beat with the carrier and with the supermodulation side frequencies which are spaced by multiples of the quench frequency from the carrier.

The adjustments of the electrode voltages relative to proper linear-mode superregeneration are critical and usually require automatic gain stabilization (a.g.s.).



1. Quench oscillator
2. Quench amplifier
3. Quench-frequency amplifier and rectifier for o-g-s control

FIG. 60. Plate-quenched superregenerative circuit with automatic gain stabilization.

beacon transponders (racons). It may be avoided by using an r-f buffer amplifier between the antenna and the detector. A superregenerative circuit may also operate at the i.f. of a superheterodyne to provide amplification and detection while the selectivity is supplied by other tuned circuits.

54. Comparison of Detectors. Detectors may be rated on the basis of the square of the output voltage in terms of the power applied to the tuned input circuit in the following order:

Type	Approximate Limits of Carrier Voltage at Tube Input
1. Superregenerative	Should exceed amplified random-noise voltage
2. Heterodyne	Same as (5) divided by r-f circuit amplification
3. Regenerative	Same as (5) divided by r-f circuit amplification
4. Grid	0.1 to 3.0 volts
5. Plate	0.2 to 10.0 volts
6. Infinite-impedance	0.5 volt to voltage level for grid current
7. Diodes	0.5 volt to level giving rated safe current in forward direction, or safe voltage in blocking direction

Types 1, 2, and 3 provide r-f amplification before detection but are not suited for systems which require a reliable undistorted replica of the original signal. Type 2 has the greatest inherent circuit selectivity since the gain at the frequency of oscillation is limited only by tube adjustments. The band width of type 3 is from 0.1 to 0.01 times that of the same tank circuit without feedback. The critical adjustments for types 2 and 3 depend on whether sensitivity, stability, or linearity is of first importance.

Types 4 and 5 have been superseded for most applications by diodes followed by a stage of audio or video amplification to restore the over-all gain. Diodes are preferred in this service since the output is linear over wide ranges of input level and percentages of modulation, and since a negative rectified voltage for a.v.c. may be obtained with simple circuit connections.

Types 4 and 7 load the tuned input circuit and increase its band width depending on its impedance relative to the tube load. Types 5 and 6 do not load the input circuit at voltage levels which do not cause grid current.

DETECTORS FOR P-M AND F-M WAVES

55. Conversion to A.M. Phase-modulated or frequency-modulated waves are detected after being converted to a-m waves. In general, a.m. is produced when an

With pulse-modulated waves where the duty cycle is small the a.g.s. may be obtained from the noise level between pulses. Either the grid bias or the amplitude of the quench voltage may be controlled. Typical diagrams are shown in Figs. 59 and 60.

A low-power wave which is pulse-modulated at the frequency of the quench is radiated from an antenna coupled directly to a superregenerative circuit. This is unobjectionable in certain applications, as in radio-

f-m wave is applied to a circuit of which the amplitude characteristic or the phase characteristic is nonuniform over the range of applied component side frequencies. In particular, when the amplitude and phase characteristics are linearly variable with frequency, a.m. proportional to the original f.m. is obtained.

A current of the form indicated in Eq. (7) applied to such a circuit produces a voltage

$$e = Z_0 I_0 [1 + Sma \cos(at - Pa)] \cos[\omega_c t + m \sin(at - Pa)] \quad (106)$$

where S = slope of impedance characteristic $\Delta Z/\Delta\omega$

P = slope of phase characteristic $\Delta\phi/\Delta\omega$

Z_0 = impedance at carrier frequency

Such an impedance is obtained approximately, over a limited frequency range, on the side of the resonance curve of a parallel-tuned circuit or near the resonant frequency of a series-tuned circuit. The resultant voltage is amplitude-modulated at the modulation frequency $a/2\pi$ to a degree Sma . The phase shift of the modulation envelope by the angle Pa and the remaining f.m. or p.m. are usually of no interest. The actual demodulation is made ordinarily in a conventional linear amplitude detector.

For f-m waves the modulation index m_f is inversely proportional to the modulating frequency $a/2\pi$ and directly proportional to the frequency deviation Δf . [See Eq. (9).]

The depth of amplitude modulation m_a obtained is

$$m_a = Sm_f a = S(\Delta\omega) \quad (107)$$

For p-m waves the modulation index (m_p) is independent of the modulating frequency. The output is then distorted unless the demodulated signal is passed subsequently through a circuit whose response is inversely proportional to the modulation frequency.

56. Balanced Frequency Detectors. A single frequency-amplitude converter of any simple type responds to spurious a.m. present in the original signal or introduced by the selective response of the i-f tuned circuits. The differentially combined outputs of two detectors, operated from converters with opposite slopes, produce a net output which is nearly linear with respect to desired frequency deviation but zero at the carrier frequency. This arrangement, when carefully tuned to the carrier, gives an output from spurious a.m. which is proportional only to the instantaneous deviation and is, therefore, masked to a great extent by the desired signal. The response to spurious a.m. may be reduced further by an amplitude limiter ahead of the detector stage, or by using a converter circuit which dynamically maintains the output level substantially constant, regardless of any changes in the input amplitude which are at an audible rate.

The most obvious balanced frequency detector or *discriminator* is a pair of circuits, one tuned slightly above and one slightly below the carrier frequency, and provided with separate rectifiers whose outputs are combined differentially. A more satisfactory type is the phase-shift discriminator shown in its more common form in Fig. 61a and in one of the many possible variations in Fig. 61b. It combines two converters of opposite slopes and is operated from a single i-f amplifier or limiter. The primary voltage plus half the secondary voltage is applied to one diode, and the primary voltage plus half the secondary voltage, with reversed polarity, to the other diode.

The shape of the output-frequency characteristic of a phase-shift discriminator is satisfactory when the secondary inductance is equal to or slightly greater than the primary inductance, when the effective Q s of the circuits are approximately equal, and when these circuits are overcoupled by an amount that depends on the available Q and the required maximum deviation.

The primary, secondary, and output voltages of the phase-shift discriminator of

Fig. 61a for an input current I at an angular frequency $2\pi(F + \Delta f)$ are approximately

$$E_1 = -2\pi FQLI \frac{1 + j2D}{4D^2 - 1 - K^2 - j4D} \tag{108}$$

$$E_2 = 2\pi FQLI \frac{jK}{4D^2 - 1 - K^2 - j4D} \tag{109}$$

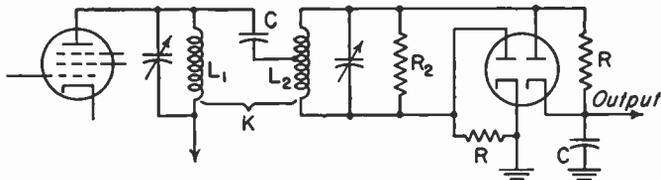
$$\text{Output} = 2\pi FQLI\eta \frac{\sqrt{1 + (2D + 0.5K)^2} - \sqrt{1 + (2D - 0.5K)^2}}{\sqrt{(4D^2 - 1 - K^2)^2 + 16D^2}} \tag{110}$$

where $D = Q(\Delta f)/F$

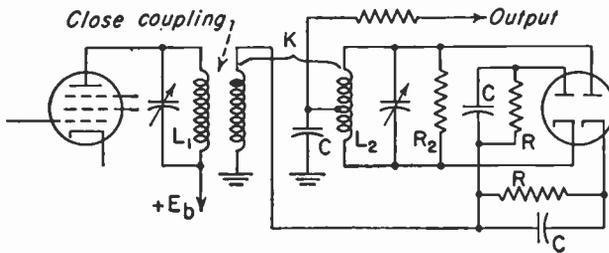
$K = kQ = Q \times$ coefficient of inductive coupling

$\eta =$ voltage rectification efficiency of diodes

The complex primary and secondary voltages, Eqs. (108) and (109), for $K = 3$ are shown in Fig. 62 for three values of Δf . The phase shifts of the primary and secondary voltages are in opposite directions as the frequency departs from the center frequency, tending to make the net output more nearly linear with frequency over a wide range.



(a) Conventional



(b) Output balanced to ground

FIG. 61. Phase-shift discriminator.

Universal curves for normalized output, expressed by the radical fraction in Eq. (110), are plotted for $K = 1, 2, 3,$ and 4 in Fig. 63. Only one-half of each curve is shown since the output curve for negative deviations is skew-symmetrical with that shown. Actual curves check closely with the calculated curves except when spurious couplings between leads or spurious capacitances to ground modify the results. (The output of the circuit of Fig. 61b has the same shape but only half the amplitude.) The value of K to be chosen, not to exceed 3.0, for maximum efficiency is such that $Q(\Delta f)/F$ for maximum deviation is as large as possible while still restricted to a linear portion of the curve. Deviation over too wide a range results in odd-harmonic distortion. Operation with the carrier considerably displaced from the frequency of zero response results in severe even-harmonic distortion.

Both diodes in parallel place a shunt load on the primary while they are in series with respect to the secondary. The resistor R_2 may be added to equalize the circuit Qs. As shown in Fig. 61a, the load resistors R are in parallel with the individual

diodes making the net effect of each diode and resistor approximately equal to $R/3$. The capacitors C are for r-f by-pass only.

A phase-shift discriminator has an advantage over separate tuned circuits in that it may be made linear over greater deviations, since when properly designed the double peaking in the primary circuit, and the opposite phase shifts in primary and secondary, compensate for some rounding of the maximum response. The tuning is also simpler since the secondary circuit is adjusted to obtain zero response at the center frequency

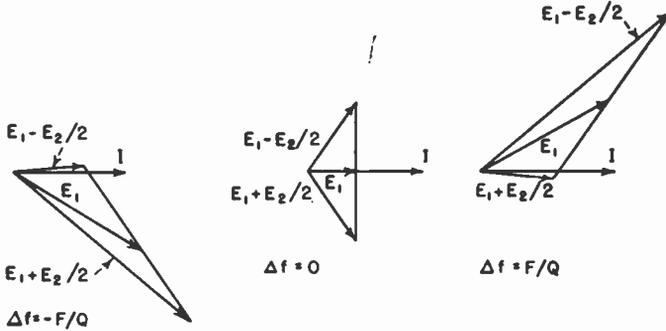


FIG. 62. Voltages in phase-shift discriminator.

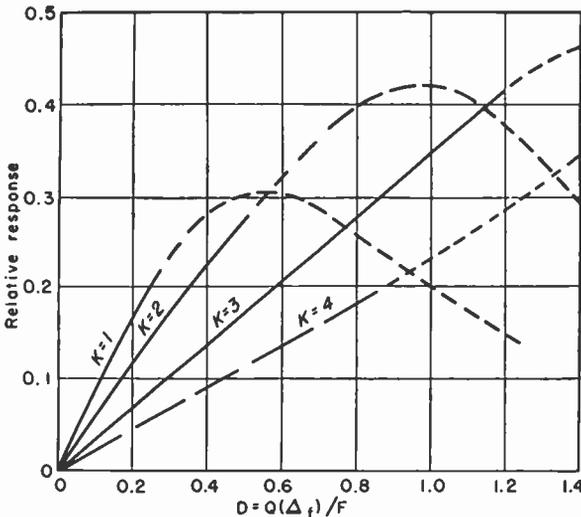


FIG. 63. Relative response of phase-shift discriminator.

while the primary circuit is adjusted to obtain symmetry of response on either side of center.

57. Ratio Detectors. A balanced discriminator very similar to that of Fig. 61b but provided with an electrolytic capacitor C_e across the two load resistors in series, as shown in Fig. 64, inherently maintains the output at a level substantially proportional to the frequency deviation and to the average i-f input regardless of a.m. Such a circuit is usually called a *ratio detector* since the ratio of the rectified voltages that appear across the two diodes is approximately equal to the ratio of the open-circuit

a-c voltages developed between the two live terminals of the discriminator windings and ground. Many other series arrangements of elements produce equivalent results. Although these circuits are relatively new, they are already in wide use.

The electrolytic capacitor is equivalent to a battery having a voltage equal to the average rectified voltage at any fixed operating level. Its voltage and the audio output voltage are substantially unaffected by rapid changes in input level, as with a.m., but follow the average input level, as with changes during tuning.

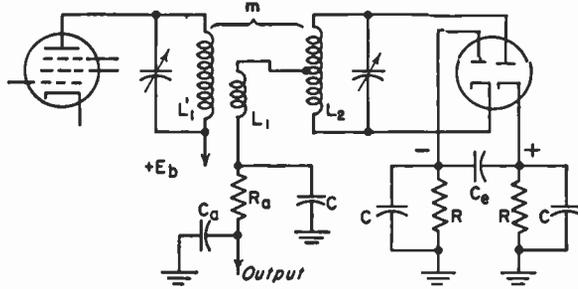


FIG. 64. Conventional ratio detector.

The same rectified current flows through both diodes and the load resistors since their d-c paths are in series. The rectified voltage across each diode is equal to its open-circuit peak a-c voltage minus the drop in the equivalent impedance of the source (see Sec. 44). The rectified current increases during outward a.m., though the division of diode voltages remains the same except for small changes in balance, owing, for example, to small differences between rectification efficiencies of the two diodes. The same is true for the balance for any downward a.m. which does not lower the instantaneous open-circuit peak a-c voltage below that across the electrolytic capacitor.

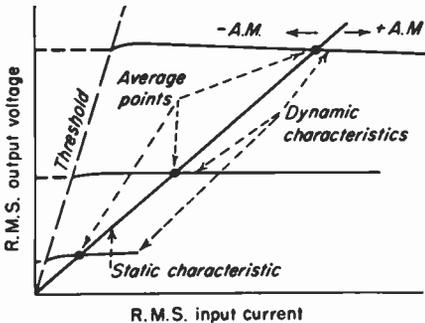


FIG. 65. Dynamic and static characteristic of ratio detector.

The open-circuit voltages and impedances of the two sources change in the same ratio during frequency deviation thereby providing differential output as in the conventional discriminator.

Effective nonresponse to a.m. in a ratio detector depends on a correlation of the load resistors with the diode efficiencies and with the impedance of the tuned circuits. This involves proper selections of the self and mutual inductances of the primary and secondary windings and of the r-f by-pass capacitance C . The rectification efficiencies cannot be more than about 50 per cent or the circuit will cut off on downward

a.m. and introduce severe distortion. The primary inductance and resistors are usually smaller than in a conventional discriminator. No complete mathematical analysis is known to exist, and designs are checked by tests for linearity of desired output and for residual a.m. at various i-f and modulation levels.

Typical characteristics showing net output voltages with variations of input current for a fixed maximum deviation are shown in Fig. 65. The static characteristics are obtained with a slow change of input while the dynamic characteristics are obtained by superimposing a.m. The slope of the dynamic characteristic is proportional to the residual a.m. It may be negative at high input levels; i.e., outward a.m. may

cause the net output to decrease. The limiting a.m. to cause diode cutoff, and the ratio of the residual to the original a.m. for about 30 per cent original a.m. are illustrated in Fig. 66.

The residual a.m. is usually divided between balanced and unbalanced effects, as shown in Fig. 67. In Fig. 67a the full line is the deviation characteristic without a.m. and the shaded area covers the limits of residual a.m. for about 50 per cent modulation at the input. The dashed lines show the limits of the envelope in the absence of amplitude reduction. The balanced and unbalanced components are shown (exaggerated) in Figs. 67b and 67c. The unbalanced effects may probably be accounted for by small changes in the effective diode input capacitances with level, and by harmonic voltage components introduced in the source impedances by harmonic components in the diode current.

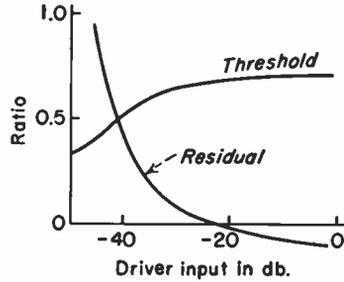


FIG. 66. Amplitude modulation in ratio-detector output.

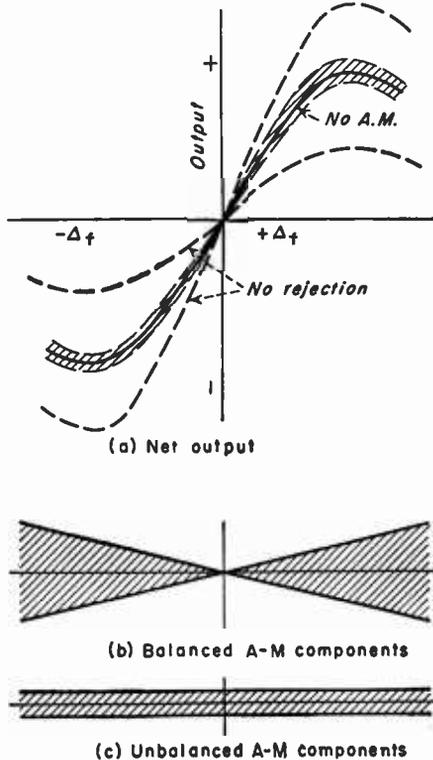


FIG. 67. Residual a.m. in ratio-detector output.

quency is lowered sufficiently to fall in step with the i.f. The control function provides more or less plate current to keep the oscillator in step as the i.f. is deviated.

¹ Bradley, W. E., Single-stage F-M Detector, *Electronics*, October, 1946, pp. 88-91.

and by harmonic voltage components introduced in the source impedances by harmonic components in the diode current.

The deemphasis time constant $R_a C_a$ may be made $75 \mu\text{sec}$ in detectors for broadcast f.m. to compensate for the specified preemphasis.

58. Single-tube Frequency Detectors. Numerous arrangements using a single multigrid tube have been suggested for providing a frequency detector which is nonamplitude responsive. The most successful incorporates the functions of an oscillator locked to the i.f. by an action in the plate circuit of a special heptode tube equivalent to that of a reactance tube controlled by the relative phase between the i-f and oscillator voltages.¹ The shortcomings of simpler locked oscillators are largely avoided (see Sec. 63).

The fundamental circuit is shown in Fig. 68a. The cathode and first grid of the heptode provide oscillation in the associated tank circuit, $L_2 C_2$. A low-Q circuit ($L_1 C_1$) tuned broadly to the i-f band and coupled loosely to the tank circuit is included in series with the plate lead. Short pulses of current at oscillator frequency appear in the plate circuit which induce a voltage in the tank circuit in quadrature therewith and proportional to their amplitude.

In the absence of an i-f voltage the oscillator operates at a frequency above the deviation band. An i-f voltage on the third grid modifies the magnitude of the plate current until the oscillator frequency

Since the d-c component of a current consisting of narrow pulses is one-half the a-c component, any changes in the required a.c. for the control of frequency are also present in the d-c component of the plate current which is used for the demodulated output. A resistor, by-passed to ground for i.f., in the plate-supply circuit provides an output impedance. A change in the amplitude of the applied i-f voltage does not change the oscillator amplitude but only modifies slightly the phase required for stable operation. Therefore, the circuit is substantially free from the effects of a.m. when the i-f voltage is sufficiently great during downward a.m. to synchronize the oscillator over the deviation range.

The output is linear with respect to frequency deviation as shown in Fig. 68b. The frequency range over which synchronism is maintained increases with the i-f

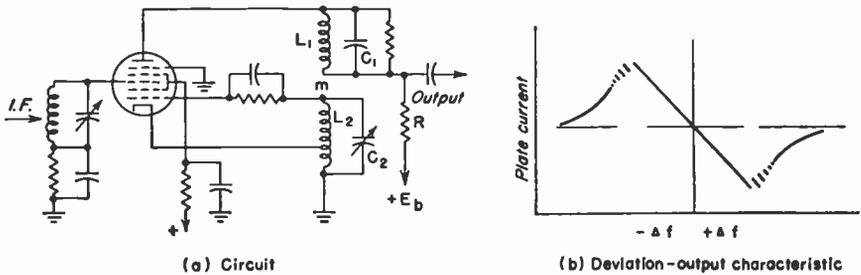


Fig. 68. Bradley f-m detector.

level as in other types of synchronized oscillators. Synchronism is lost when the deviation is excessive.

59. Harmonic Distortion. The output of a discriminator with a linear deviation characteristic follows faithfully the instantaneous frequency deviations applied to it. Curvature in the characteristic produces harmonic and intermodulation distortion. The latter is much more disturbing to the ear in audio work, but systems are rated more easily in terms of simple harmonic distortion. High-order harmonics are the most disturbing and may be weighted in terms of the second harmonic. An example of weighting is given in the following table:

Harmonic	Relative Weight
1	1
2	2
3	4
4	10
5	

The allowable weighted distortion and residual a.m. (see Sec. 57) depend on the class of service but usually should not exceed the ratios indicated in the following table:

Fidelity	Residual a.m.	Weighted harmonics
Medium.....	0.35	0.10
Good.....	0.2	0.05
High.....	0.1	0.02

A curved characteristic, skew-symmetrical with respect to zero, produces only odd-harmonic distortion while an unsymmetrical one gives even-harmonic distortion as well. The detection distortion from deviation over a slowly curving characteristic, such as for $K = 1$ in Fig. 63, is mainly the third harmonic, while that from a curve such as for $K = 3$, also contains considerable fifth and seventh. The total percentage

in each case is approximately one-third the percentage departure from linearity at the extremes of deviation. For example, for $K = 3$ and $D = 1.4$ in Fig. 63, there will be approximately 2 per cent distortion; if D does not exceed 1.2, there should be no detection distortion.

The over-all effective discriminator characteristic may be narrower than for the discriminator alone, owing to the contribution of the selective circuits as illustrated in Fig. 69. Exact relations depend on the degree of response to a.m. and a-v-c action. Mistuning the receiver to the peaks of the over-all curve practically eliminates the response to the fundamental of the signal frequency. The output is then all even-harmonic distortion. Mistuning the carrier beyond the peak to $\Delta f = A$ on the curve often provides as great response as when properly tuned although with considerably more distortion. This is especially noticeable when a fixed-threshold amplitude limiter, such as a grid limiter, is used. The side responses are reduced considerably with dynamic limiting to provide an over-all operating characteristic similar to that of a ratio detector. The proper tuning point is then identified easily as giving greater volume.

Distortion is introduced by too great deviation in the i-f pass band even though the discriminator is perfect. This originates in the nonuniform phase slope on the sides of the selectivity curve. It is inadvisable to design for deviation over the skirts of an over-all selectivity curve where the response is more than 6 or 10 db below that at the center frequency, as measured with linear amplification.

60. Phase Detectors. It is necessary to use a frequency detector for p-m waves, followed by a corrective network, when the modulation index is greater than about 0.5 radian, to avoid distortion of the demodulated output.

When the modulation index is small, a p-m wave may be combined with an auxiliary carrier synchronized 90 deg out of phase with the modulated carrier. The composite wave is then amplitude-modulated by the reverse of the process mentioned in Sec. 9.

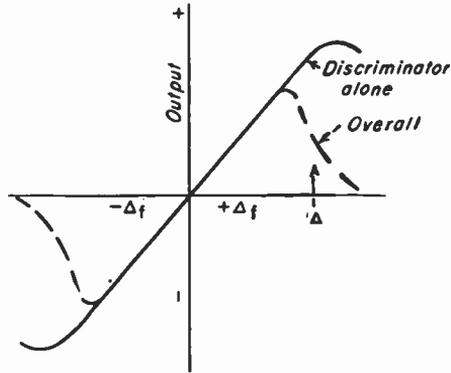


FIG. 69. Contribution of i-f selectivity to discriminator performance.

MISCELLANEOUS APPLICATIONS

61. Grid-bias Amplitude Limiters. An overloaded class C amplifier with grid-leak bias may be used to develop substantially constant fundamental-frequency plate current over a range of about 10 or 15 to 1 in input voltage.

The usual circuit is shown in Fig. 70. The plate circuit may be connected directly to a discriminator and in any case is tuned to the desired frequency, rejecting the higher harmonics. The tube is operated ordinarily at a fixed low screen voltage to prevent excessive current in the absence of a signal and to place the knee of the input-output characteristic at a low voltage.

The tube operates as a class A amplifier below the knee. The grid-cathode circuit acts as a diode whose instantaneous current may be expressed by Eq. (45). Resistance R is selected to develop bias which allows the plate current flow during smaller portions of each cycle as the input voltage is increased. The

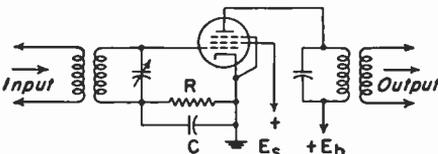


FIG. 70. Grid-bias limiter.

slope of the input-output curve is determined almost entirely by the product $Rh\sqrt{E_c}$, where R is the grid resistance, h is the coefficient in Eq. (45) for grid current, and E_c is the negative grid voltage necessary to cut off the tube when connected as a pentode at the screen voltage selected.

Normalized steady-state input-output curves are shown for three values of grid resistance in Fig. 71. The abscissas are in terms of E_a/E_c , where E_a is the peak value of the applied i-f voltage and E_c is defined above. The ordinates are in terms of peak fundamental-frequency plate current divided by $g_m E_c$. The value of $Rh\sqrt{E_c}$ for substantially flat limiting is between 35 and 40.

The grid current loads the input circuit as with a diode, and test curves should be taken in terms of the input voltage at the grid of the preceding tube. The capacitance is made as small as is consistent with over-all gain so that the grid bias may follow rapid changes in the amplitude of the input.

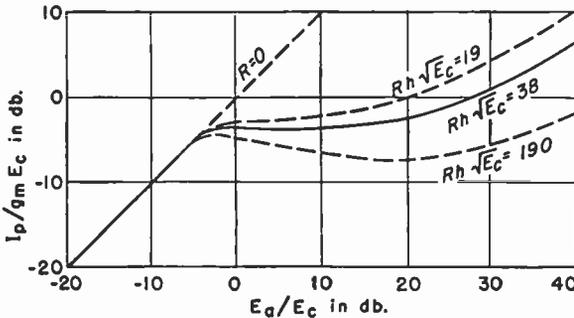


Fig. 71. Grid-bias limiter characteristics.

Grid-bias limiters are not effective when operated below the knee of the input-output characteristic. The lower limit of input signal at which limiting begins is fixed by the knee. Consequently, a grid-bias limiter will handle greater percentages of downward a.m. at high average carrier levels.

Limiters are often used in cascade when the input range over which the output is constant does not meet the requirements. The output voltage of the first limiter is then adjusted to fall just above the knee of the second.

62. Dynamic Diode Limiters. A vacuum-tube or germanium-crystal diode fed from a tuned circuit and with a low load resistance shunted by an electrolytic capacitor may be used as a dynamic limiter for the fundamental component of the a-c circuit voltage. The over-all dynamic and static operating characteristics are quite similar to those shown in Fig. 65 for the ratio detector except that the output curves always rise slowly with increase of input current.

The peak circuit voltage during a.m. is equal to the capacitor voltage as fixed by the average input divided by the rectification efficiency of the diode for the modulated current. The residual a.m. is always positive. The limit of downward a.m. that can be handled without diode cutoff is determined by the modulated peak voltage, which, when multiplied by the impedance of the source, just equals the capacitor voltage.

63. Synchronized Oscillators. An oscillator operated at a submultiple of the i.f. may be synchronized and operates with little response to a.m. The carrier frequency and deviation are then reduced in the same ratio, and the variations in the amplitude of the output are considerably less than in the applied i.f. Operation at the i.f. is inadvisable since a very small amount of feedback to the input of the i-f amplifier modifies the phase shift and makes the operation unstable.

The ability to control an oscillator over a range of frequencies is improved when the phase shift per cycle of frequency deviation is small. This factor may be reduced by coupling a low- Q circuit to the oscillator tank circuit, tuned to the same frequency, with less than critical coupling.¹

¹Carnahan, C. W., and H. P. Kalmus, Synchronized Oscillators as F-m Receiver Limiters, *Electronics*, August, 1944, pp. 108-111, 332-342.

Advantages of synchronized oscillators used in reducing the effects of a.m. are (1) an increase in effective selectivity equal to about two stages of i-f amplifier; (2) stable gain and a considerable reduction in over-all a.m.; (3) a quieting sensitivity inversely proportional to the deviation of the synchronizing signal. Synchronized oscillators have not been used widely for this purpose because of the following disadvantages: (1) the reduction of a.m. is inferior to that of other simpler systems; (2) a high input voltage must be supplied for synchronization; (3) there is a tendency to pull out of synchronism if improperly tuned; (4) subharmonic operation is lost if the synchronizing voltage is too high; (5) the input circuits must be fed from a low-impedance source and they load the previous i-f amplifier and reduce its gain and selectivity; (6) the linearity of operation is dependent on the impedance of the discriminator connected to the output; (7) synchronization may be stable over too narrow an i-f band. A solution of the last two items has been discussed in connection with an oscillator operated at one-fifth the i.f.¹

64. Diode Clippers. Diodes may be used as limiters or *clippers* of video waves either in series or in parallel with the load. Figures 72a and b show two typical examples of many possible arrangements.

In Fig. 72a, when the resistances of the source and load are equal, the first diode passes current when the input voltage is more positive than $-E/2$, and the second diode when the input voltage is less than $+E$. No current reaches the load outside of these limits except through capacitance coupling between diode elements. It is desirable that the resistances be large compared to the forward resistances of the diodes. The circuit is suitable at low frequencies of input signal where the time constant of the circuit capacitances with the resistances is small compared to the period of the wave.

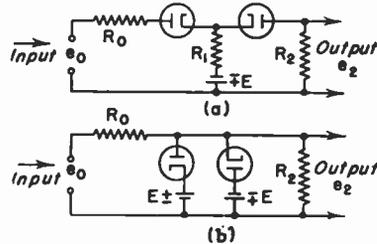


FIG. 72. Types of diode clippers.

In Fig. 72b each diode is nonconducting as long as the output voltage is less than $\pm E$. For larger values of voltage (\pm), one diode or the other shunts the load with its low forward resistance. For effective limiting, the resistances of the source and load must be considerably larger than the diode resistance.

Stages may be used in cascade, alternated with video amplifying stages, to obtain nearly rectangular l-f wave forms.

In both arrangements shown in Fig. 72 a small bias voltage may be needed for symmetrical clipping on both half waves to neutralize the zero-current diode voltage. When clipping in one polarity only is needed, one diode may be omitted.

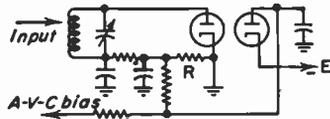


FIG. 73. Threshold limiter for quiet automatic volume control.

develop a voltage drop in R equal to $-E$, the second diode is biased beyond cutoff while the first functions as a normal detector without bias. At signal levels at which each diode carries some current, there is peak clipping when the signal is fully modulated.

66. Frequency Multipliers. Any tube with a nonlinear grid-plate characteristic contains harmonics of the input frequency in its plate circuit. The output circuit may be tuned to the desired multiple of the original frequency and the other components by-passed to ground. Frequency multipliers are used mainly for constant-

¹ Beers, G. L., A Frequency-dividing Locked-in Oscillator Frequency-modulation Receiver, *Proc. IRE*, December, 1944, pp. 730-737.

amplitude or telegraph waves. An expansion of the power-series expression for plate current shows that the degree of modulation of an a-m wave is increased by each multiplication, and intolerable distortion results.

When an a-m voltage wave

$$e = E(1 + m \cos at) \cos \omega t$$

is applied to a square-law tube, the third term of Eq. (48) yields a modulated current wave at twice the original carrier frequency

$$i = \frac{A_2 E^2}{2} \left(1 + 2m \cos at + \frac{m^2}{2} + \frac{m^2}{2} \cos 2at \right) \cos 2\omega t \quad (111)$$

The new degree of modulation m' for the fundamental is

$$m' = \frac{4m}{2 + m^2} \quad (112)$$

In addition there is a second harmonic modulation

$$m'' = \frac{m^2}{2 + m^2} \quad (113)$$

When a p-m or an f-m voltage wave

$$e = E \cos (\omega_c t + m \sin at)$$

is applied to a square-law tube, the third term in Eq. (48) yields a modulated current wave at twice the original carrier frequency

$$i = \frac{A_2 E^2}{2} \cos (2\omega_c t + 2m \sin at) \quad (114)$$

The modulation index m and the carrier frequency are multiplied by the same figure in the successive frequency multiplications of p-m or f-m waves.

Class A multipliers, in which the plate current is never cut off, are inefficient. More economical use of the power capability of the tube is realized in class C service where the grid-bias and input voltage are adjusted for maximum output. The expansion of the power series Eq. (48) then contains many frequency terms of more than twice the original carrier frequency.

The proper operating conditions are best determined by test. The bias is approximately that for plate-current cutoff in the absence of signal in doublers and somewhat greater than this for triplers. A slight amount of feedback at the harmonic frequency increases the output and the plate efficiency.

Higher harmonics than the third may be selected for laboratory work although the available power is limited.

Two tubes may be used in push pull with increased efficiency. The outputs are connected in parallel for doublers and in push pull for triplers.

CHAPTER 12

RADIO-WAVE PROPAGATION

BY DALE POLLACK¹

1. Classification into Frequency Ranges. For purposes of analysis it is convenient to divide radio waves into bands of frequencies within which propagation effects are similar. Any such classification must, in part, be arbitrary, since changes in the properties of waves with frequency are not sharply defined and are dependent upon time. In this chapter the classification of Table 1 will be followed in general, but minor changes in the dividing frequency between ranges will be made when desirable. The column Approximate Useful Communication Radius is quite approximate, the actual radius being dependent upon the power, time, earth properties, and other conditions.

2. General Characteristics of Radio Waves. The signal intercepted by a radio-receiver antenna may have been propagated by any of three waves from the transmitter: the *surface wave*, which travels along the earth's surface, the *sky wave*, which leaves the antenna at relatively high angles and reaches the receiver by reflection from ionized layers above the earth; or the *space wave*, which is received more or less directly through space from the transmitting antenna. The combination of the surface wave and space wave is commonly called the *ground wave*. The received signal is, in the general case, made up of components of all three types, but it is often convenient to investigate transmission by each of the modes independently. Frequently one or the other predominates, facilitating such independent analysis.

The *surface wave* is attenuated by losses in the earth and falls off at distances from the transmitter, the exact manner in which it decreases being dependent upon the frequency and the conductivity and dielectric constant of the earth. The surface wave is useful only for medium-distance communication at low frequencies and for short distances at medium frequencies. At higher frequencies it can be employed only for local communication.

Table 1. Classification into Frequency Ranges

Frequency range, Mc	Nomenclature	Approximate useful communication radius	
		Ground wave, miles	Sky wave (or, above 30 Mc, tropospheric wave), miles
0.03- 0.3	Low frequency	0-1,000	500-8,000
0.3 - 3	Medium frequency	0- 100	100-1,500
3 - 30	High frequency	0- 30	100-8,000
30 - 300	Very high frequency	0- 50	50- 150
300 - 3,000	Ultrahigh frequency	0- 50	?
3,000 -30,000	Super high frequency	0- 50	?

The *sky wave* may be employed for communication over greater distances since it travels through the atmosphere, in which the attenuation is relatively small. It can be refracted and reflected by the ionized layers in the upper atmosphere (the iono-

¹ Consulting Radio Engineer, New London, Conn.

sphere) and by the earth. Thus the wave may return to earth at distances remote from the transmitter. All frequencies, up to the very high, may be reflected or refracted to earth by the ionosphere, and their sky waves may, therefore, be useful.

If the ground wave is to render good service at a distant receiving point, it must be strong compared with (1) the noise and interference level and (2) other waves. The limit to the useful service range of the surface wave may be fixed by one or both of these factors. The surface wave may be increased with respect to the noise and interference level either by increasing the transmitter power or by employing a directional antenna; with respect to the sky wave, the only effective aid is the use of an antenna system in which high-angle radiation is minimized. In the region in which the sky wave and the surface wave are nearly equal in magnitude (within, perhaps, 2 to 1 of each other) fading, particularly selective fading, is excessive. At locations beyond the region in which the surface and sky waves are of similar magnitude, the sky wave predominates and is more useful for communication.

The space wave is usually made up of two components, as shown in Fig. 1: the *direct wave* from the transmitting antenna and the *reflected wave* from the surface of the

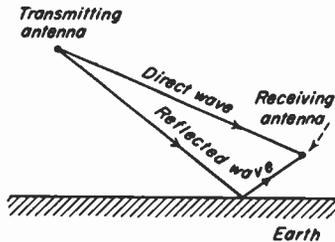


FIG. 1. Relation between direct and reflected waves.

earth. The space wave exists only when the two antennas are a great distance (in wavelengths) from the earth's surface, usually only for frequencies above 30 Mc. Above 30 Mc, radio waves are very rapidly attenuated over the earth's surface, and their sky wave is not normally returned to earth by the ionosphere. Consequently, the transmission properties of such waves must depend upon the direct ray from the antenna and upon reflections of the direct ray from the earth's surface, as assisted by diffraction around the curved surfaces of the earth and refraction in the lower atmosphere. Their usefulness is limited to relatively short distances although greater than the optical line-of-sight limitation which is too frequently assumed. As the frequency increases, however, the effects of refraction in the lower atmosphere become more and more erratic in character, and longer range communication may be expected to become less reliable. Consequently, communication by means of waves higher than 300 Mc is limited to paths that are only a little longer than optical.

3. Propagation of the Surface Wave. The distance for which the surface wave is useful decreases as the frequency is raised. The surface wave of low frequencies is useful up to medium distances, and in the broadcast band its usefulness is limited to short distances. Above the broadcast band it can be used only for local transmission.

Sommerfeld¹ has computed the propagation of waves over a plane earth, *i.e.*, for distances short enough that the earth's curvature may be neglected. The results may be further simplified if the dielectric constant of the earth is neglected, which is true within an error smaller than 2 to 1, if the frequency is less than

$$f_c = \frac{7 \times 10^{20} \sigma}{\epsilon} \quad (1)$$

where f_c = frequency below which dielectric constant may be neglected, cycles

σ = soil conductivity, mks units, mhos per m

ϵ = dielectric constant, mks units (air = 10^{11})

Typical values of the soil constants and f_c are shown in Table 2. Data on soil conductivity throughout the United States are shown in Fig. 2. In using such a map, it must be remembered that it shows only gross effects. Local conditions, such as the presence of stone buildings in cities, will cause local variations which the map cannot indicate.

¹ Sommerfeld, A., The Propagation of Waves in Wireless Telegraphy, *Ann. Phys.*, **28**, 665-736, Mar. 16, 1909; **31**, 1135-1153, Dec. 11, 1926.

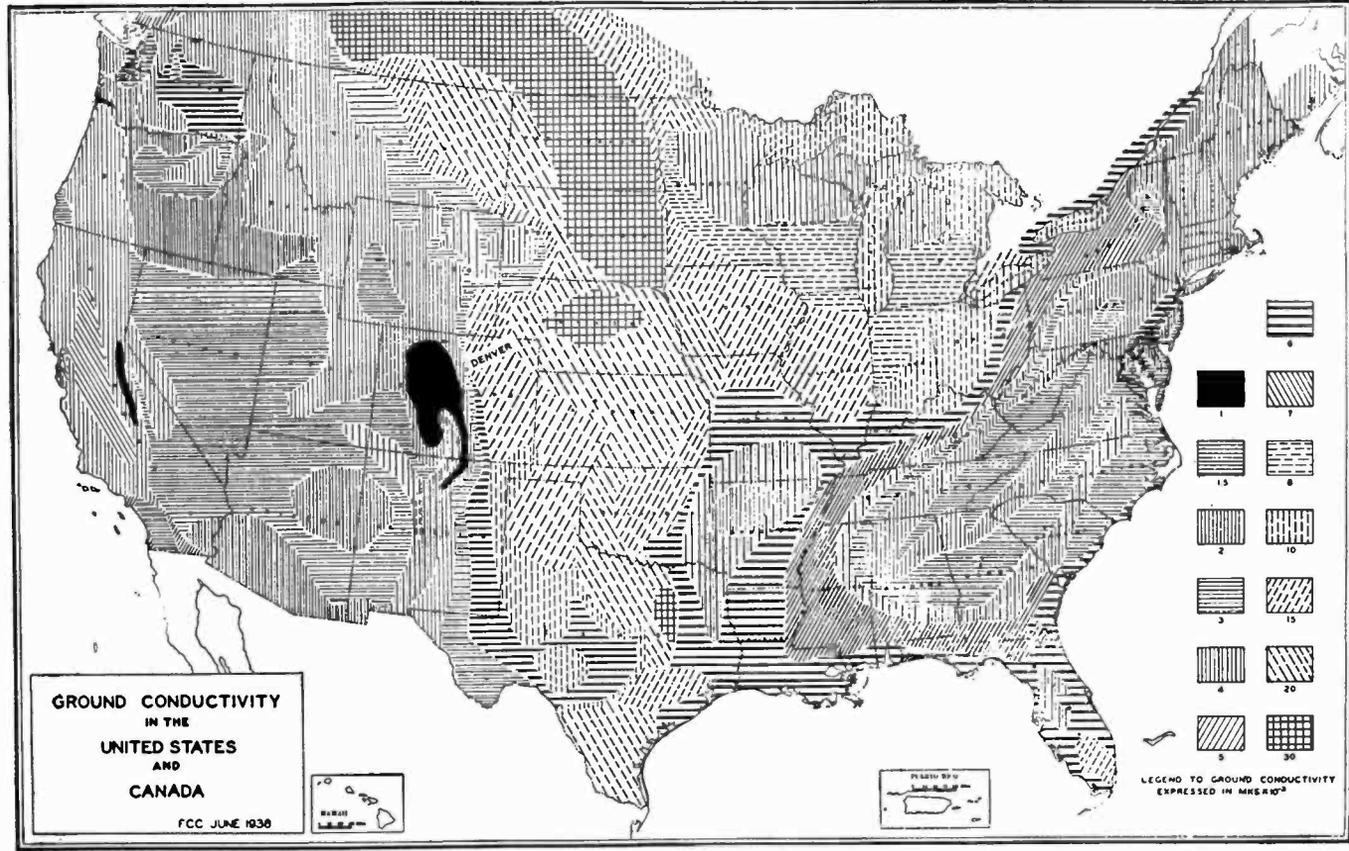


FIG. 2. Earth characteristics of North America.

Table 2

	σ (mks)	ϵ (mks)	f_c (cycles)
Dry, sandy, or rocky ground.....	10^{-3}	5×10^{11}	1.4×10^4
Eastern and far western U.S. (approx).....	3×10^{-3}		
Average ground, central U.S. and Europe (approx).....	10×10^{-3}		
Moist ground.....	30×10^{-3}	30×10^{11}	7×10^4
Salt water.....	4	80×10^{11}	350×10^4

From the above assumptions, the surface-wave field intensity, according to van der Pol's¹ simplification of Sommerfeld's work, is given by

$$\epsilon = \frac{K \sqrt{P} A}{d} \quad \text{volts/m} \quad (2)$$

where K = antenna constant (10 for $\lambda/4$ antenna or 15 for $\lambda/2$ antenna)

d = distance, m

P = power radiated, watts

$$A = \frac{2 + 0.3 \times 10^{-11} \rho}{2 + 10^{-11} \rho + 0.6 \times 10^{-22} \rho} \quad (\text{Sommerfeld's "reduction factor"})$$

$$\rho = \frac{5.8 \times 10^{-19} f^2 d}{\sigma} \quad (\text{Sommerfeld's "numerical distance"})$$

f = frequency, cycles

σ = soil conductivity, mks (mhos/m)

When large numbers of such data must be computed, the slide rule of J. F. Morrison² is convenient. In practical computations of wave propagation, account must be taken of changes in the soil conductivity within the transmission range. Procedures for accomplishing this are outlined by P. P. Eckersley³ and by H. L. Kirke.⁴

More recent analyses have extended propagation calculations to account for the earth's curvature. Curves giving the results of such calculations will be found in the Report of Committee on Radio Wave Propagation.⁵ Two such sets of curves for two typical earth conductivities (those of average ground and sea water, respectively) are shown in Figs. 3 and 4. For radiated powers different from 1 kw, the field intensities of these figures should be multiplied by \sqrt{P} , where P is the radiated power in kilowatts. For other antenna structures the field should be multiplied by the appropriate factor. For a $\lambda/2$ antenna, for example, this factor is 1.4.

4. Propagation of the Sky Wave. Energy which leaves the antenna at angles greater than zero with the earth's surface constitutes the sky wave, in contrast to the surface and space waves which leave the antenna at zero or lower angles. The sky wave travels through the atmosphere until it reaches the ionized layers 100 to 500 km above the earth, called the "ionosphere" or the Kennelly-Heaviside layers. It may then be bent back to the earth immediately, or it may travel in the ionosphere for some distance before being returned to earth, or it may pass through the ionosphere and never return to earth. Long-distance radio communication is almost always accomplished by means of the portion of the sky wave reflected to earth. After the wave returns to earth, it may be reflected from the earth's surface into the ionosphere once more and to the earth at a more distant point.

¹ van der Pol, B., Propagation of Electric Waves, *Z. Hochfrequenz*, **37**, 152-156, April, 1931. For excellent discussion of surface-wave propagation, see K. A. Norton, The Propagation of Radio Waves over the Surface of the Earth and in the Upper Atmosphere, Part I (Ground Wave), *Proc. IRE*, **24**, 1367-1387, October, 1936.

² For sale by Keuffel & Esser Co., Hoboken, N.J.

³ The Calculation of the Service Area of Broadcast Stations, *Proc. IRE*, **18**, 1160-1193, July, 1930.

⁴ Report on Two Methods of Calculating Ground Wave Field Strength of a Transmitter over a Composite Land and Sea Path, *Brit. Broadcasting Co., Rpt. KO55, Serial No. 1943/16*, July 7, 1943.

⁵ *Proc. IRE*, **26**, 1193-1234, October, 1938.

The medium through which the surface wave is propagated, the surface of the earth, changes little with time. Sky-wave propagation, however, depends upon the ionosphere characteristics, which vary widely with time. The engineer is normally interested more in the effect of those ionosphere characteristics on the propagation of waves than in the characteristics themselves. These effects are outlined here, while the ionosphere characteristics are considered in Sec. 10.

Sky-wave propagation is conveniently subdivided into low, medium, and high frequencies. The sky wave at low frequencies is usually considerably stronger at night than during the day. The *day* field for over-water transmission may be estimated from the modified Austin-Cohen formula:

$$\epsilon = 0.377 \times 10^6 \frac{hI}{\lambda d} \sqrt{\frac{\theta}{\sin \theta}}$$

$$\epsilon = \frac{45 \times 10^{-4} d}{\lambda^{0.6}} \text{ volts/m} \quad (3)$$

- where I = antenna current, amp
- d = distance, m
- θ = angle at center of earth subtended by transmission path in radians
- h = effective height, m
- λ = wavelength, m

The quasi-maximum field (the field intensity which is exceeded only 5 per cent of the time) for a propagation path completely in darkness is given by Fig. 5.¹ For long distances a distinction must be made between transmission paths near the magnetic pole (corresponding to a north-south or south-north transmission, such as between Europe and South America, or between North and South America) and transmission paths far from the earth's magnetic pole (corresponding in general to east and west transmission, such as between northern United States and northern and central Europe, or between northern and central Europe and Siberia).¹ The median value of the field (the field exceeded 50 per cent of the time) is about 35 per cent of the quasi-maximum value. For powers other than

¹ *Ibid.*

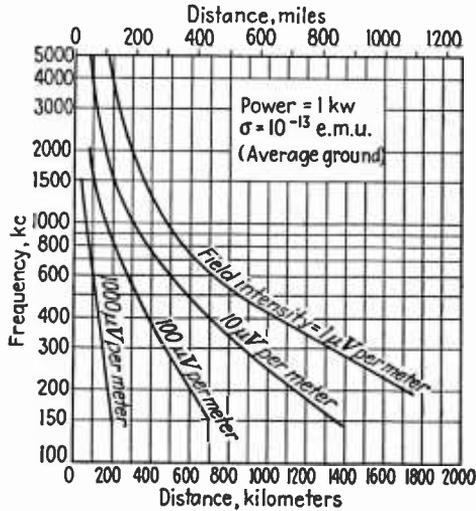


FIG. 3. Ground-wave propagation over average ground; 1 kw, vertical $\lambda/4$ antenna. (Report of Committee on Radio Wave Propagation, Proc. IRE, 26, 1193, October, 1938.)

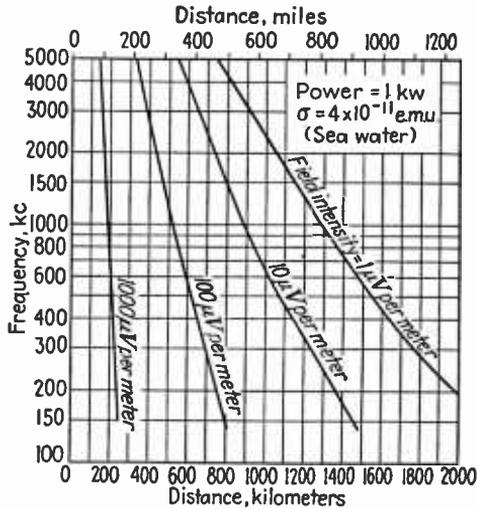


FIG. 4. Ground-wave propagation over sea water; 1 kw radiated power, vertical $\lambda/4$ antenna. (Report of Committee on Radio Wave Propagation, Proc. IRE, 26, 1193, October, 1938.)

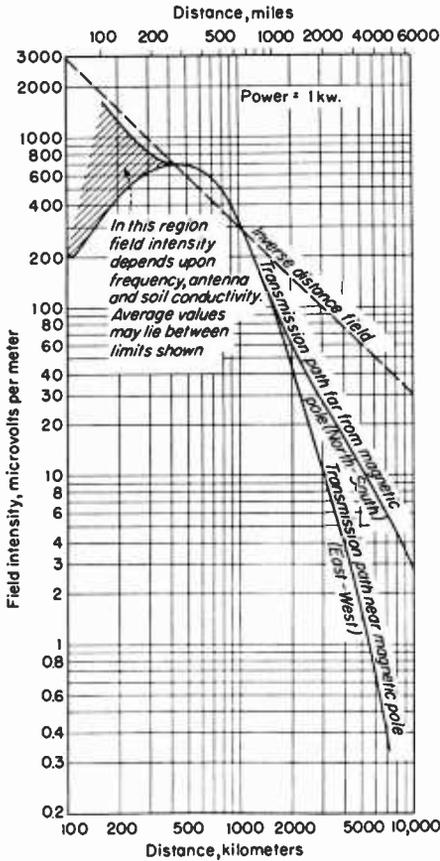


FIG. 5. Sky-wave propagation, quasimaximum field intensity for frequencies up to 1,600 kc. (Report of Committee on Radio Wave Propagation, Proc. IRE, 26, 1193, October, 1938.)

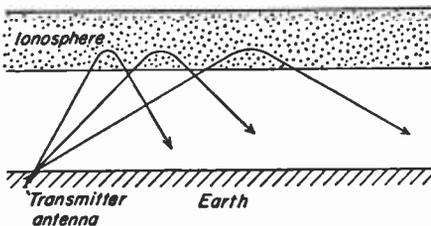


FIG. 6. Probable paths traversed by skywave at low frequencies. A single ionized layer is shown.

1 kw multiply the field intensity by \sqrt{P} where P is in kilowatts. P should also include the antenna power gain if a directional antenna is employed.

When part of the transmission path is in twilight, l-f propagation suffers. For long-distance communication the field strength is likely to be lower than during either night or day transmission and its magnitude is not easily predicted.¹

At medium frequencies, the sky wave does not return to earth during the day. Sky-wave propagation for such frequencies, therefore, need be considered only for nighttime communication. Such propagation to about 1,600 kc is represented in Fig. 5. Such considerations determine long-distance nighttime broadcast transmission, "secondary coverage," and are discussed in greater detail in the chapter on Radio Broadcasting.

Above about 1,600 kc, sky-wave propagation is a complex phenomenon. For each of the ionosphere layers there exists a critical frequency below which radiation from any angle from the antenna is returned to earth. Above this critical frequency, high-angle radiation passes through the ionized layer, while radiation from lower angles is still returned to earth. This is illustrated by Figs. 6 and 7. In Fig. 6, for a frequency smaller than the critical value, the entire sky wave is returned to earth. In Fig. 7, however, the frequency is greater than critical for nearby receivers, and the high-angle radiation is never returned to earth. The distance between the transmitter and the point at which the radiation of highest angle returns to earth is the skip distance for that frequency. Conversely, for this smallest skip distance, the corresponding frequency is called the maximum usable frequency (MUF). In general, the higher the frequency, the greater the skip distance.

Within the skip distance, the signal strength is usually too weak to be

¹ Espenchied, L., C. N. Anderson, and A. Bailey, Trans-Atlantic Radio Telephone Transmission, Proc. IRE, 14, 7-57, February, 1926.

useful. This explains the necessity for employing frequencies lower than the maximum usable values. In traversing nonionized air h-f waves suffer little attenuation, other than that resulting from spreading of the wave front. In passing through ionized air, however, the attenuation is greater. An operating frequency should be chosen, therefore, for which as little of the path as possible is in the ionosphere. This will be the case if a frequency only slightly lower than the maximum usable one is employed.

The MUF is dependent upon the time of day (or longitude), month, year, and latitude, as well as upon the distance between transmitter and receiver. The curves of Figs. 8a and b give the values over which the MUFs may be expected to range during summer and winter, respectively. The curves are from data taken between 1933 and 1939, but they may be considered as typical for corresponding portions of the sunspot cycle.¹ Upper limits of the curves are for sunspot maximums (1937-1938 and 1948-1949); lower limits are for sunspot minimums (1932-1934 and 1944-1945). The time (noon and midnight are shown in the figures) and latitude (39° north) refer to the place at which the

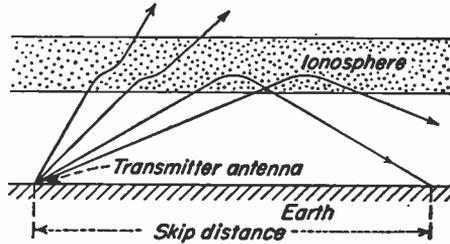


FIG. 7. Probable paths at high frequencies.

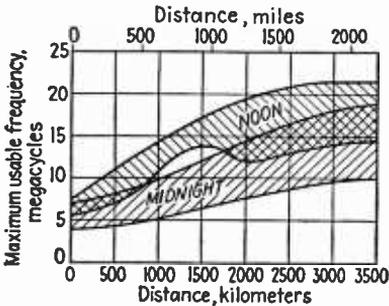


FIG. 8a. Average maximum usable frequencies during summer at latitude 39°N. Time refers to place where wave is reflected. The upper edge at each range is for the sunspot maximum, 1938-1939, the lower edge for the sunspot minimum, 1933-1934. (Bureau of Standards Ionosphere Reports, Monthly in Proc. IRE; Smith, Gilliland, and Kirby, Nat. Bur. Standards, J. Research, 21, 835, December, 1938; Gilliland, Kirby, Smith, and Reyer, Proc. IRE, 26, 1347, November, 1938.)

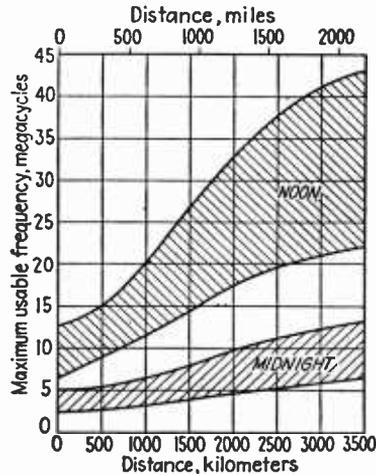


FIG. 8b. Average maximum usable frequencies during winter, at latitude 39°N. Time refers to place at which wave is reflected. The upper edge of each range is for the sunspot maximum, 1938-1939. The lower edge is for the sunspot minimum, 1933-1934. (Bureau of Standards Ionosphere Reports, Monthly in Proc. IRE; Smith, Gilliland, and Kirby, Nat. Bur. Standards, J. Research, 21, 835, December, 1938; Gilliland, Kirby, Smith, and Reyer, Proc. IRE, 26, 1347, November, 1938.)

ionosphere reflection takes place, usually halfway between transmitter and receiver

¹ Smith, N., T. R. Gilliland, and S. S. Kirby, Trends of Characteristics of the Ionosphere for Half a Sunspot Cycle, J. Research Nat. Bur. Standards, 21, 835-846, December, 1938; Gilliland, T. R., S. S. Kirby, N. Smith, and S. E. Reyer, Maximum Usable Frequencies for Radio Sky-wave Transmission, 1933-1937, Proc. IRE, 26, 1347-1359, November, 1938.

for transmission paths less than 3,500 km long. The curves represent average conditions only; during the relatively infrequent periods of ionosphere disturbances, the MUFs may be considerably changed; above 50 Mc, they have been observed for short periods of time during sunspot maximums. While the measurements were made at latitude 39° north, the results may probably be used with insignificant error between 30 and 50° north.

Systematic measurements of ionosphere characteristics have been carried out by the U.S. Bureau of Standards and elsewhere, since the early 1930's. Results of these measurements have appeared as follows:

May, 1933, to April, 1934: Gilliland, *Proc. IRE*, **23**, 1076-1101, September, 1935.
 May, 1934, to December, 1936: Gilliland *et al.*, *Proc. IRE*, **25**, 823-840, July, 1937.
 January, 1937, to May, 1937: Gilliland *et al.*, *Proc. IRE*, **25**, 1174-1184, September, 1937.
 June, 1937, to October, 1941: Gilliland *et al.*, monthly in *Proc. IRE*, September, 1937, through December, 1941.

Publication of these reports was interrupted during the war, but they have since been resumed:

U.S. Bureau of Standards, Central Propagation Laboratory, *Basic Radio Propagation Predictions*, a monthly publication.

From a knowledge of the ionosphere characteristics over the earth, it is now possible to compute the propagation of h-f waves.¹ During the past few years, sufficient ionosphere data have been accumulated to permit such calculations to be made. Propagation maps, for frequencies of 8.6 and 18.8 Mc, calculated entirely from ionosphere data are given in the Report of Committee on Radio Wave Propagation.²

More complete methods of prediction are given in the Bureau of Standards publication referred to above.

5. Propagation of Very High and Ultrahigh Frequencies. Very-high-frequency waves, as intercepted by a receiver antenna within the horizon, are made up of two components: one received directly from the transmitter antenna, the second reflected by the earth. When the distance between transmitter and receiver is small enough so that the earth's curvature may be neglected, the received field strength may be computed from

$$\varepsilon = \frac{4.2 \times 10^{-9} h_t h_r f \varepsilon_0}{d^2} \quad \text{volts/m} \quad (4)$$

where h_t and h_r = heights of transmitting and receiving antennas, respectively, m
 ε_0 = field intensity 1 m from transmitter, volts per meter, in the direction of maximum field strength (for a $\lambda/2$ dipole, $\varepsilon_0 = 7 \sqrt{P}$)
 P = transmitter power, watts
 d = distance between transmitter and receiver, m
 f = frequency, cycles

This equation is derived by the simple addition of the direct and reflected waves, assuming that the earth is a perfect plane reflector and that both antennas are at least several wavelengths above the earth's surface. This is, of course, a very much simplified treatment, but it is often adequate for estimates of field strength well within the horizon.

6. Propagation beyond Line of Sight. It was originally believed that radio waves above 30 Mc in frequency were nearly optical in character and, therefore, could be propagated very little, if any, beyond the horizon. This is erroneous, and satisfactory

¹ An early attempt at such computation was by S. Namba and T. Tsukada, A Method of Calculation of Field Strengths in High Frequency Radio Transmission, *Proc. IRE*, **21**, 1003-1028, July, 1933. More recent and more satisfactory methods are presented in the following papers: Smith, N., Application of Vertical-incidence Ionosphere Measurements to Oblique-incidence Radio Transmission, *J. Research Nat. Bur. Standards*, **20**, 683-705, May, 1938; Millington, G., The Relation between Ionosphere Transmission Phenomena at Oblique Incidence and Those at Vertical Incidence, *Proc. Phys. Soc. (London)*, **50**, 801-825, September, 1938.

² *Loc. cit.*

v-h-f communication is practical over distances two or more times beyond this optical line of sight. Such propagation is made possible by diffraction of the waves around the curvature of the earth and by refraction and reflection in the lower atmosphere. Diffraction around the earth is a definite, reproducible effect, the unknowns usually being the earth's conductivity and dielectric constant. Atmospheric effects are variable, however, depending upon the temperature, barometric pressure, and the water content of the air within the propagation path.

The theory of v-h-f propagation has recently been improved to take account of diffraction of waves around the curved surface of an earth of finite conductivity.¹ The results, from the Report of Committee on Radio Wave Propagation² are plotted in Figs. 9 through 12 for frequencies of 50 and 150 Mc for typical land and salt-water conditions. The curves are for vertical polarization, but apply without much error to horizontal polarization as well. The curves take account of diffraction around the earth's curvature, but not of refraction and reflec-

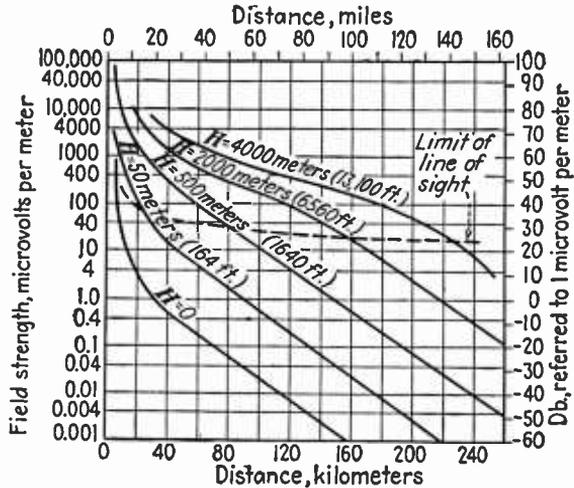


FIG. 9. Ultra-high-frequency propagation over average land ($\epsilon = 5, \sigma = 10^{-13}$ emu) at 50 Mc. Antenna current equals that produced by 1 kw in antenna at surface of earth. (Report of Committee on Radio Wave Propagation, Proc. IRE, 26, 1193, October, 1938.)

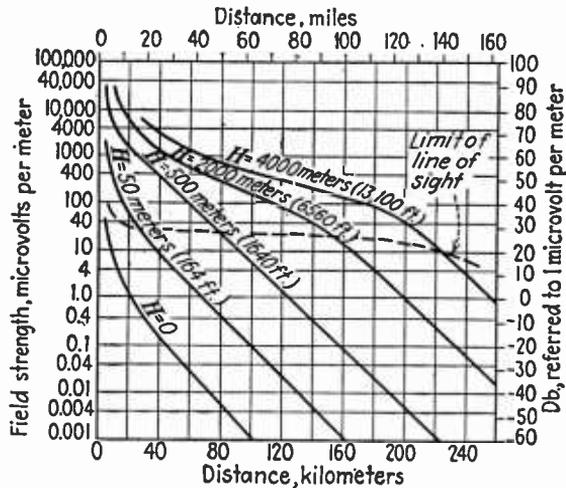


FIG. 10. Ultra-high-frequency propagation over land ($\epsilon = 5, \sigma = 10^{-13}$ emu) at 150 Mc. (Report of Committee on Radio Wave Propagation, Proc. IRE, 26, 1193, October, 1938.)

¹ Watson, G. N., The Diffraction of Electric Waves by the Earth, Proc. Roy. Soc. (London), 95, 83-99, Oct. 7, 1918, and 546-563, July, 1919; Eckersley, T. L., Ultra-short-wave Refraction and Diffraction, J. IEE, 80, 286-304, March, 1937; van der Pol, B., and H. Bremmer, The Diffraction of Electromagnetic Waves from an Electric Point Source round a Finitely Conducting Sphere with Applications to Radio Telegraphy and the Theory of the Rainbow, Phil. Mag., 24, Part I, 141-176, July, 1937; and Part II, 825-864, November, 1937; Norton, K. A., The Calculation of Ground Wave Field Intensity over a Finitely Conducting Spherical Earth, Proc. IRE, 29, 623-639, December, 1941.

² Loc. cit.

tion in the atmosphere. The current in the antenna is that for 1 kw in the antenna when the antenna is at the earth's surface. It is assumed that the antenna current is held constant as the antenna is raised by correcting the transmitter power. The correction required is nominal, however.

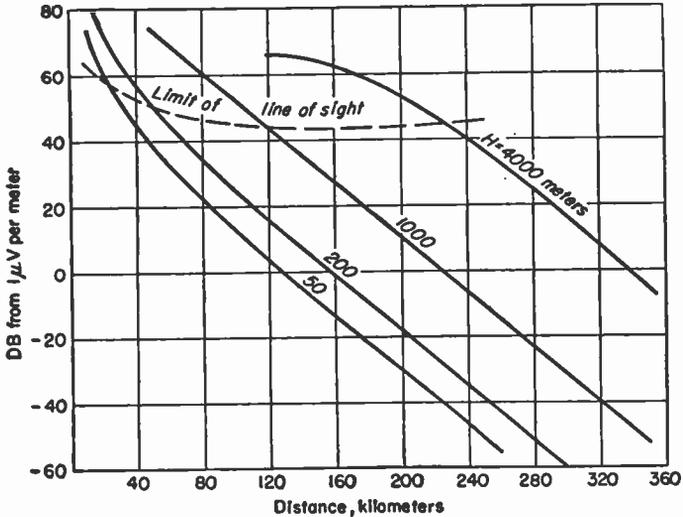


FIG. 11. U-h-f propagation over sea at 50 Mc.

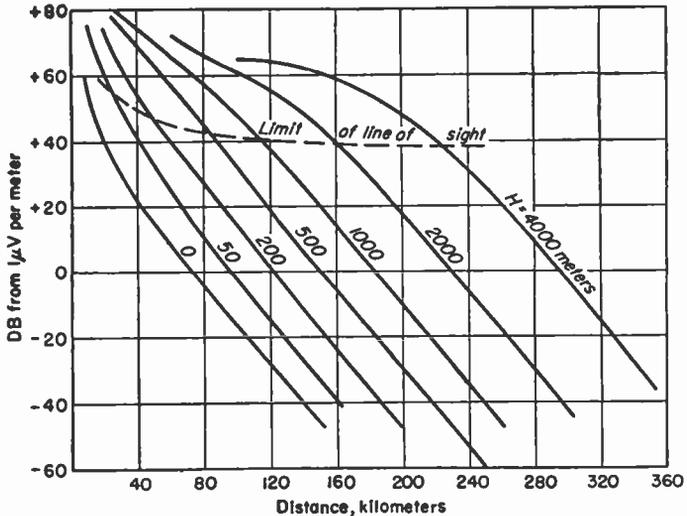


FIG. 12. U-h-f propagation over sea at 150 Mc.

The curves are plotted for several values of H , which represents either the receiver or transmitter antenna height, assuming the other antenna height to be zero. If both antennas are above the surface of the earth, a correction must be applied. For distances beyond the horizon, correction curves are plotted in Figs. 13 and 14, each curve being associated with the conditions of one of the previous figures.

7. Line-of-sight Propagation. Optical line-of-sight distances may be computed from

$$D_{\text{optical}} = 1.23(\sqrt{h_t} + \sqrt{h_r}) \quad \text{miles} \quad (5)$$

where the antenna heights are in feet. Optical line-of-sight distances are plotted in

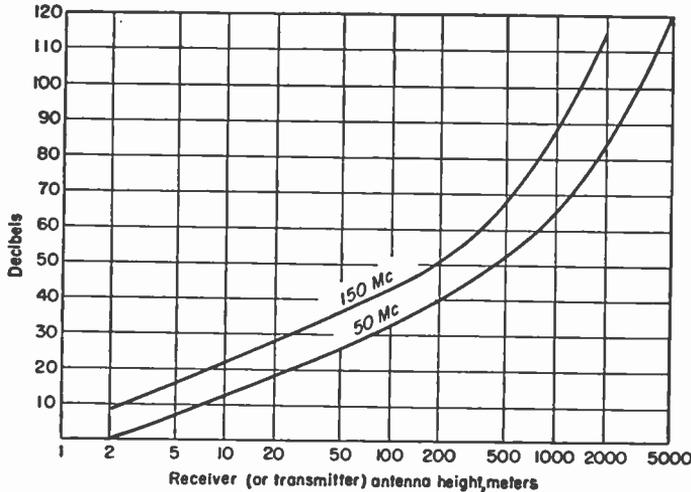


Fig. 13. Correction curves for u-h-f propagation for distances beyond the horizon over land.

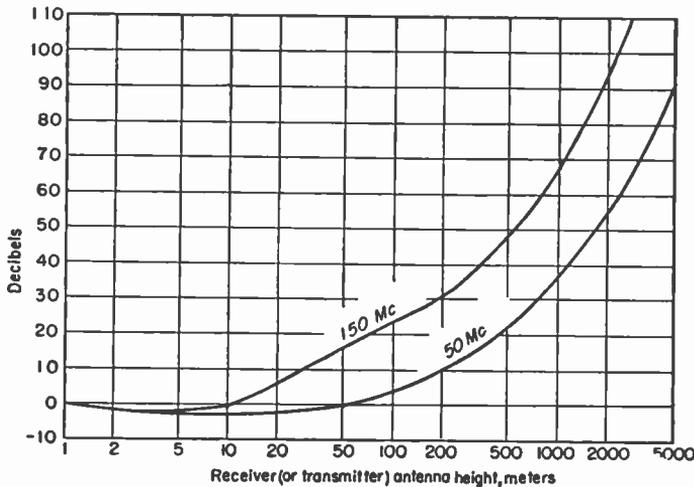


Fig. 14. Correction curves for u-h-f propagation for distances beyond the horizon over sea.

Fig. 15 and are also indicated in Figs. 9 through 12. It is of interest to note that the field strength at the line-of-sight distance is nearly independent of the antenna height. For 1 kw radiated it is approximately 30 μv per m over land and 100 μv per m over sea. While attenuation is more rapid beyond the line-of-sight distance, nevertheless reliable v-h-f communication is perfectly practical to distances far beyond the optical horizon, as is evident from the previous propagation data. However, the rate of

attenuation beyond the horizon increases with frequency, and, for u-h-f waves, transmission is not possible very far beyond the line of sight. Within the horizon the attenuation may be expressed, roughly, by an inverse square relationship, Eq. (4); beyond the horizon the attenuation is better expressed by an exponential equation of the form

$$\epsilon = ke^{-\alpha d} \quad (6)$$

in which k and α are constants independent of the distance. If ϵ is measured in micro-

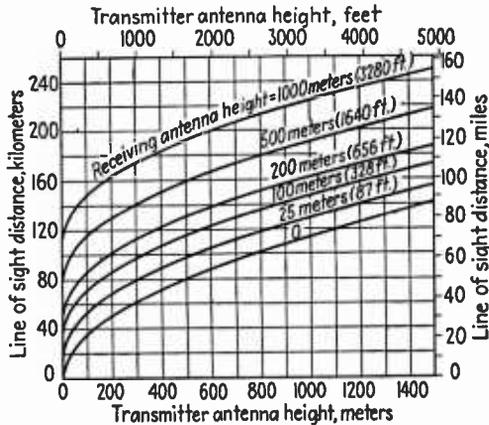


FIG. 15. Length of line-of-sight path for various antenna heights.

be considered to represent probable minimum signal strengths. Refraction has two effects. In addition to increasing the average received field strength, refraction causes the field to fade as the density and water-vapor content of the air in the transmission path fluctuate. Most of the time the signal will fall between two limits: the diffraction value as previously given as a minimum and the free space field (which varies inversely as the distance) as a maximum. For small percentages of the time, the field will vary beyond these extremes.

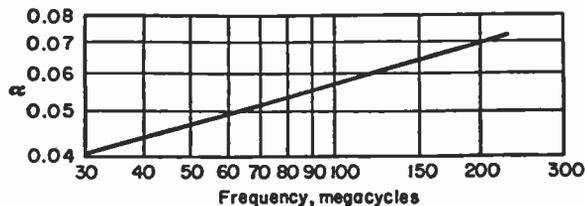


FIG. 16. Attenuation factor for line-of-sight transmission.

It is only recently that the effects of changes in the index of refraction have been understood. It is now possible to explain them and to correlate them with weather and atmospheric measurements.¹ It is expected that in the near future it will be

¹ Smith-Rose, R. L., and J. S. McPetrie, Ultra-short Waves: Refraction in the Lower Atmosphere, *Wireless Engr.*, 11, 3-11, January, 1934; Englund, C. R., A. B. Crawford, and W. W. Mumford, Further Results of a Study of Ultra-short-wave Transmission Phenomena, *Bell System Tech. J.*, 14, 369-387, July, 1935; Hull, R. A., Air-wave Bending of Ultra-high-frequency Waves, *QST*, 21, 16-18, May, 1937; Friend, A. W., and K. C. Colwell, The Heights of the Reflecting Regions in the Troposphere, *Proc. IRE*, 27, 626-634, October, 1939; Smyth, J. B., and L. G. Trolese, Propagation of Radio Waves in the Lower Troposphere, *Proc. IRE*, 35, 1198-1202, November, 1947.

possible to predict probable fading ranges, if not actual field intensities, at a particular time and place.

Atmospheric refraction is associated almost entirely with the air near the surface of the earth, up to about 1,000 m. The case most frequently considered in the literature¹ is the so-called *standard atmosphere* for which the index of refraction is inversely proportional to elevation. The choice of the word "standard" is unfortunate, since the situation represented is not prevalent or standard in any sense. The increased field as a result of "standard" refraction may be accounted for by assuming the radius of the earth (and line-of-sight distances) to be increased by approximately $\frac{1}{3}$. The result, as far as the curves (Figs. 9 through 12) are concerned, is a reduction in the slope of the linear portions, and of α in Eq. (6), by the factor 0.825 and an increase in the "radio" line of sight by $\frac{1}{3}$.

In calculations of the effects of various distributions of the index of refraction, it is convenient to plot a "modified" index, given by

$$M = n - 3.86h - 1 \quad (7)$$

where n = index of refraction

h = height, m

Plots of three typical variations in the modified index of refraction are shown in Fig. 17. *A* shows the so-called "standard" case with a uniform decrease in index of refraction

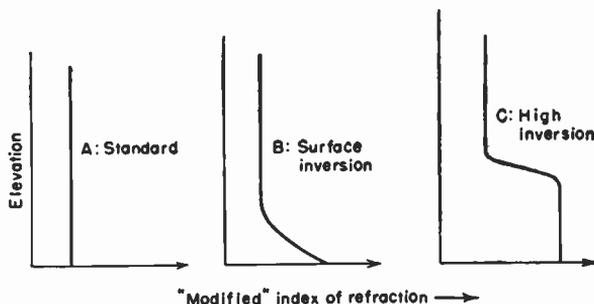


FIG. 17. Plots of three typical variations in the modified index of refraction.

with altitude. *B* is a typical inversion at the surface of the earth, caused by high water-vapor content or low temperature near the earth's surface. *C* illustrates a sharp change in the index at the boundary between two air masses, with the upper mass having a lower vapor content and being warmer than the surface mass. V-h-f waves are reflected from discontinuities in the index of refraction such as those shown at *B* and *C*. As the discontinuity increases in height, the strength of signals at the upper end of the v-h-f band is reduced, and such signals fade violently. For low-altitude inversions, on the other hand, high fields at the upper end of the v-h-f band may be anticipated. Although these maximum fields may exceed those at lower frequencies, the fading range for frequencies at the lower end of the band is reduced by these low-altitude inversions, and the lower frequencies—say, 40 to 60 Mc—are generally to be preferred for uninterrupted communication beyond the line of sight, irrespective of the type of inversion prevailing at a particular time. These conclusions apply to transmission beyond the line of sight, which is largely determined by such tropospheric weather conditions. Additional data on v-h-f fading are given in the next section.

As a result of these effects, beyond the horizon observed fields are generally stronger in spring, summer, and fall than in winter. Furthermore, they increase in intensity at sunset and then decrease again 5 to 8 hr later. Variations from these general patterns are, of course, the rule as local weather conditions change.²

¹ Norton, *loc. cit.*

² Burrows, C. R., A. Decino, and L. E. Hunt, Ultra-short-wave Propagation over Land, *Proc. IRE*

The exposition of v-h-f propagation outlined above omits one important factor: no consideration is given to irregularities in the surface of the earth. A hill between transmitter and receiver will, of course, attenuate the signal, giving the "shadow loss." The magnitude of this loss may be determined from Fig. 18.¹ These curves are computed from the diffraction of a wave over a sharp edge. Although they give more accurate results than the smooth-earth theory does, they do not account for reflections

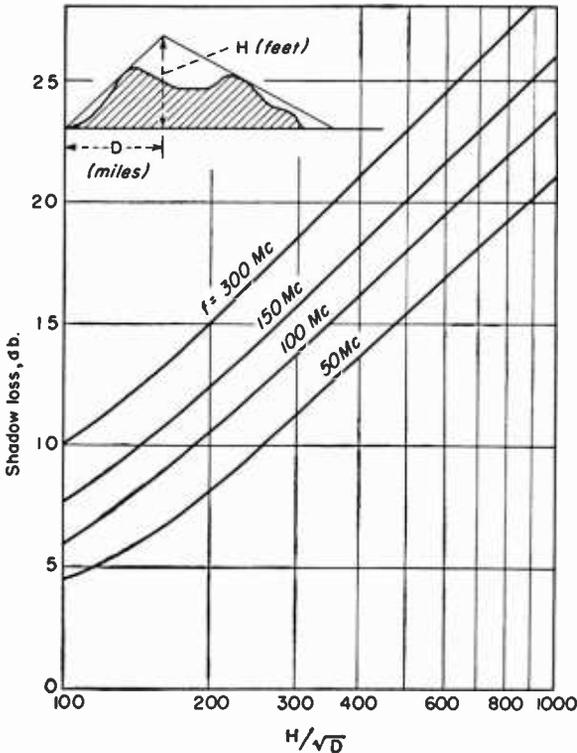


FIG. 18. Shadow loss for v-h-f transmission.

small changes in the length of one of the transmission paths may have a considerable effect on the strength of the received signal. Such fading, therefore, is usually more rapid and the range of the fading greater than that in single-path transmission. In addition, as will be indicated below, the frequency and phase characteristics of the medium may be imperfect and the wave is distorted during transmission.

When the propagation is dependent upon frequency, the fading is called *selective*. The effects of *selective fading* on reception will be shown by means of the following simplified analysis:

from nearby hills and buildings. In some locations, such as between tall buildings in cities, the prediction of signal strength is difficult.²

9. Fading. Fading, according to the standards of the IRE, "is the variation in intensity of radio signals resulting from changes in the transmission medium." Fading may conveniently be divided into two types. In the first the radio wave reaches the receiver antenna over a single path. In the second the received wave is made up of two or more components traveling over paths of different lengths.

In *single-path* transmission the received field depends directly upon the properties of the transmission medium. In such transmission, fading is likely to consist of slow variations, and the distortion in the received wave is negligible.

In *multipath* transmission

²³, 1507-1535, December, 1935; Stability of Two-meter Waves, *Proc. IRE*, **26**, 516-528, May, 1938; de Mars, P. A., and T. A. Wright, Practical Considerations of FM Coverage, *FM & Television*, **7**, Part I, 21-26, April 1947, and Part II, 31-34, May 1947.

¹ Adapted from "Propagation Curves," Division 15, National Defense Research Committee, October, 1944.

² Burrows, C. R., L. E. Hunt, and A. Decino, Ultra-short-waves in Urban Territory, *Elec. Eng.*, **45**, 115-124, January, 1935.

The voltage at the receiver antenna is assumed to be made up of two components, received over different paths:

$$v = V_1 \sin \omega t + V_2 \sin (\omega t - \psi) \tag{8}$$

where v = received voltage

V_1 = amplitude of one component

V_2 = amplitude of second component

ω = angular velocity of signal = $2\pi f$

ψ = phase angle by which second component is delayed with respect to first

This equation may be manipulated to give

$$v = V_2 \sqrt{2 \frac{V_1}{V_2} \cos \psi + \left(\frac{V_1}{V_2}\right)^2 + 1} \sin \left(\omega t - \tan^{-1} \frac{\sin \psi}{\frac{V_1}{V_2} + \cos \psi} \right) \tag{9}$$

If the difference in lengths between the two paths is Δs , then the phase angle ψ is

$$\psi = \frac{\omega \Delta s}{23 \times 10^8} \tag{10}$$

where the velocity of propagation has been assumed to be 3×10^8 m per sec. Equations (9) and (10) show that the amplitude, represented by the square root factor in Eq. (9), and the phase angle, represented by the arc-tangent term of the received signal, both depend upon the frequency of the signal and the ratio of the magnitudes of the two received components.

The amplitude factor, the square root in Eq. (9), is plotted in Fig. 19, while the angle factor, the argument of the sine in Eq. (9), is shown in Fig. 20. If an a-m signal is transmitted, the phase and amplitude relationships between carrier and side bands may be seriously disturbed if the distance between successive minimum points in Fig. 19 is comparable in magnitude to the width of the frequency band transmitted. This is particularly true when the signals transmitted over the two paths are of similar magnitude, i.e., V_1 approximately equal to V_2 . If the difference between the path lengths is known, the delay time and the frequency difference between successive minimums of Fig. 19 may be calculated from the following relations:

$$\Delta f = \frac{1}{\Delta t} = \frac{3 \times 10^8}{\Delta s} \tag{11}$$

where Δf = frequency difference between successive minimums of Fig. 19, cps

Δt = time by which one ray is delayed compared to the other, sec

Δs = difference in path lengths, m

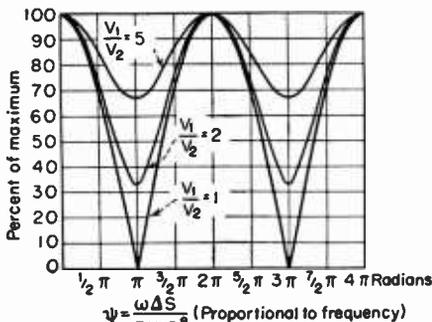


FIG. 19. Variation in amplitude against frequency in selective fading. V_1 and V_2 are the amplitudes of the two received components.

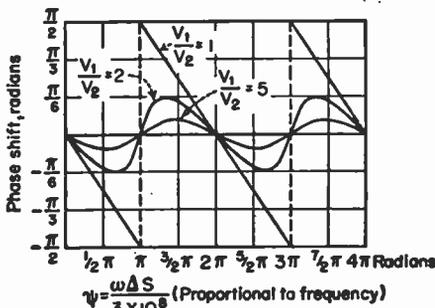


FIG. 20. Variation in phase against frequency in selective fading. V_1 and V_2 are as in Fig. 19.

The *rapidity* of fading in multipath transmission depends upon the rate at which the path lengths change. A change in the difference between path lengths of $\lambda/2$ may bring about a change from a minimum to a maximum in a fade.

The selectivity or nonselectivity of fading depends upon the numerical value of Δf and the width of the band transmitted. When one of the transmission paths is by way of the ionosphere and another is the surface wave, the difference in path lengths is of the order of hundreds of kilometers, making Δf of the order of kilocycles. This is a type of selective fading common in the broadcast band at night, and it limits the primary service range of high power broadcast stations to something between 100 and 200 km, at which distances the sky wave and ground wave have similar amplitudes. The presence of small amounts of undesired f.m. causes a particularly pernicious type of audible distortion when selective fading occurs.¹ With modern broadcast transmitters, however, the amount of residual f.m. is small. The periodicity of broadcast band fading is usually large, of the order of minutes. Such slow changes may be accommodated by conventional a-v-c circuits, provided that the minimum of the fade does not drop below the noise level and that the distortion resulting from selective fading is not excessive.

At high frequencies fading is in general more severe. Reception over a multiplicity of paths is common. The surface wave is rarely a factor in transmission, and differences between path lengths are often longer, giving values of Δf as low as a few hundred cycles.² Short-wave fading has many periods. As mentioned in connection with h-f propagation, seasonal, yearly, diurnal variations take place, but in addition much shorter periods exist, some less than $\frac{1}{10}$ sec, which are not readily accommodated with conventional a-v-c arrangements.

High-frequency fading may be minimized through *diversity reception*. If two receiving antennas are spaced several wavelengths apart, it has been observed that the signals picked up do not fade in synchronism. Accordingly, if several antennas, normally three, spaced approximately 10λ apart, are employed, sufficient output is almost always available from at least one of the antennas to provide a useful signal.³ Distortion resulting from selective fading is usually worse on the poorer signals. Diversity radiotelephone systems, therefore, are commonly arranged to provide nearly all the l-f output voltage from the strongest signal automatically. The use of single-side-band signals also assists in avoiding distortion of this kind.

The effects of multipath transmission may be further avoided by the use of receiving antennas, directional in both the horizontal and vertical planes and aimed to pick up the strongest component of the signal.⁴ Antennas whose directivity is under the control of the operator have been developed for this purpose.⁵

As noted in the paragraphs on v-h-f propagation, the field intensity calculated from the diffraction of waves around the curved surface of the earth is increased by refraction and reflection in the lower atmosphere. Changes in the density and vapor content of the air, therefore, cause changes in the propagation medium. Such changes result in variations in signal level, which become especially severe beyond the line of sight.

Figure 21 shows the fading ranges which may be expected at 100 Mc, plotted in terms of the ratio of the actual transmission distance to the line-of-sight distance. The reference level—0 db—is the signal level which would be calculated based upon

¹ Bown, R., DeL. K. Martin, and R. K. Potter, Some Studies in Radio Broadcast Transmission, *Proc. IRE*, 14, 57-132, February, 1926; Eckersley, T. L., Frequency Modulation and Distortion, *Exptl. Wireless and Wireless Engr.*, 7, 482-487, September, 1930.

² Potter, R. K., Transmission Characteristics of a Short-wave Telephone Circuit, *Proc. IRE*, 18, 581-648, April, 1930.

³ Beverage, H. H., and H. O. Peterson, Diversity Receiving Systems of RCA Communications, Inc., for Radio Telegraphy, *Proc. IRE*, 19, 531-561, April, 1931; Peterson, H. O., H. H. Beverage, and J. B. Moore, Diversity Telephone Receiving System of RCA Communications, Inc., *Proc. IRE*, 19, 562-584, April, 1931.

⁴ Friis, H. T., C. B. Feldman, and W. M. Sharpless, The Determination of the Direction of Arrival of Short Radio Waves, *Proc. IRE*, 22, 47-78, January, 1934.

⁵ Friis, H. T., and C. B. Feldman, A Multiple Unit Steerable Antenna for Shortwave Reception, *Proc. IRE*, 26, 841-917, July, 1937.

a uniform gradient in the index of refraction of the atmosphere, the so-called "standard" atmospheric case. For high antennas (the extreme case plotted is for an antenna about 2,000 m above the earth) the fading range is greater than for low antennas (100 m is the extreme plotted). The curves are for signals which are exceeded 1 and 99 per cent of the time.

Figure 21 is for 100 Mc. At lower frequencies the fading ranges will be smaller and conversely. The data have been assembled from several sources.¹

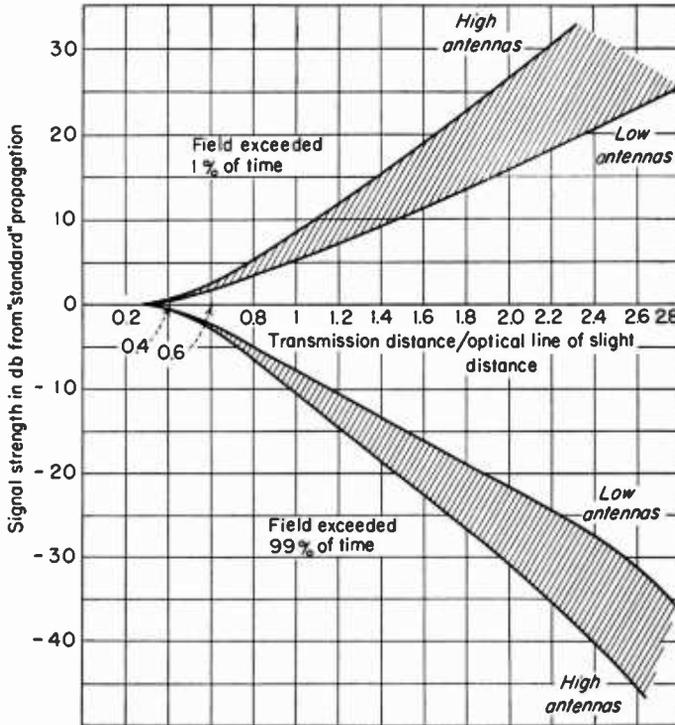


FIG. 21. Fading ranges at 100 Mc.

10. Ionosphere Characteristics.² All long-distance radio communication takes place through refraction of waves in the ionized layers above the earth. Ionosphere research has been accelerated during recent years, and knowledge of ionosphere characteristics has increased accordingly.

The ionization is apparently caused by ultraviolet radiation from the sun, and it varies with the 11-year sunspot cycle, the time of day, the time of year, and the longitude. At the earth's surface the ion density is very small, increasing to maximum

¹ MacLean, K. G., and G. S. Wickizer, Notes on the Random Fading of 50-megacycle Signals over Non-optical Paths, *Proc. IRE*, 27, 501-506, August, 1939; de Mars and Wright, *loc. cit.*

² For a general introduction to the subject see P. O. Pederson, "The Propagation of Radio Waves," published by G. E. C. Gad, Copenhagen (in English), or K. K. Darrow, The Ionosphere, *Elec. Eng.*, 59, 272-283, July, 1940. A historical survey and résumé of information is contained in S. S. Kirby, L. V. Berkner, and D. M. Stuart, Studies of the Ionosphere and Their Application to Radio Transmission, *Proc. IRE*, 22, 481-521, April, 1934, and is brought up to date in Report of Commission II, Radio Wave Propagation, International Scientific Radio Union, *Proc. IRE*, 27, 645-649, October, 1939; and in T. R. Gilliland, S. S. Kirby, N. Smith, and S. E. Reymer, Characteristics of the Ionosphere and Their Application to Radio Transmission, *Proc. IRE*, 25, 823-840, July, 1937.

values at altitudes between 100 and 500 km. While the ions are distributed continuously in the atmosphere, the concentrations vary and several maximums are reached at various altitudes. The regions near these maximums are called "layers." For each such layer there is a critical frequency above which an electromagnetic wave, directed vertically, will not be returned to earth. Ionosphere characteristics are frequently studied from measurements of critical frequencies.¹

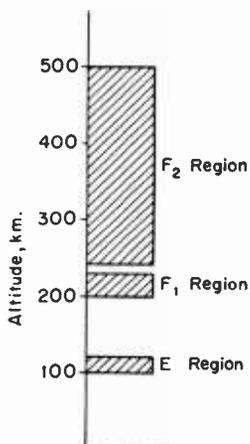


FIG. 22. Heights over which ionosphere layers range.

usually constant in height at 100 to 120 km. However, it is not always detectable at night, especially during high sunspot activity.

Within the *E* layer are sometimes found ionized "clouds" with sharply defined boundaries. These clouds move about in unpredictable fashion, thus being called *sporadic E* variations. Such sporadic *E* ionization occasionally will reflect v-h-f waves, to 60 or 70 Mc, giving very long-distance transmission on such frequencies. Such occurrences are quite rare, however.

The *F* layer, which exists only at night, exhibits little diurnal or seasonal change in height. Its height does, however, change during the sunspot cycle. The *F* layer divides into the *F*₁ and *F*₂ layers during the day. The *F*₂ layer varies symmetrically about noon and is considerably higher in summer than in winter. The *F*₂ layer appears to be the daytime continuation of the *F* layer, while the *F*₁ layer is nonexistent during the night.

A wave directed toward the ionosphere is split into two components: one in the direction of the earth's magnetic field, the other at right angles to it. The two rays are called the ordinary and the extraordinary rays and, since the magnetic field acts upon them differently, their propagation in the ionosphere is

The three ionosphere regions of greatest importance, as far as radio communication is concerned, are denoted the *E*, *F*₁, *F*₂ layers. The range of altitudes over which these layers may vary is shown in Fig. 22. At night the *F*₁ and *F*₂ layers merge, forming the *F* region. It is also possible that a *C* layer, lower than the *E*, and a *G* layer, higher than the two *F* layers, exist.²

The ion density in the upper atmosphere probably varies in some such manner as shown in Fig. 23. Only the virtual heights of the maximum points and the apparent densities at these heights are known. For intermediate altitudes it is usually assumed that the ion density goes through minimum points as shown.

The variations in the heights of the layers are shown in Figs. 24 and 25, which give typical summer and winter diurnal variations during a sunspot minimum (1933-1934) and a sunspot maximum (1938-1939).³ The *E* layer is

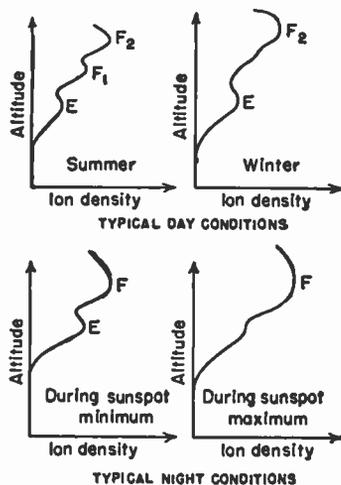


FIG. 23. Probable variation in atmospheric ion density with altitude.

¹ Gilliland, T. R., Multifrequency Ionosphere Recording and Its Significance, *Proc. IRE*, **23**, 1076-1101, September, 1935.

² Kirby, S. S., and E. B. Judson, Recent Studies of the Ionosphere, *Proc. IRE*, **23**, 733-751, July 1935; Cowell, R. C., and A. W. Friend, The Lower Ionosphere, *Phys. Rev.*, **50**, 632-635, Oct. 1, 1936.

³ Smith, Gilliland, and Kirby, *loc. cit.*; Gilliland, Kirby, Smith, and Reymer, *loc. cit.*

different.¹ One consequence of this is that two critical frequencies exist for each ionized layer, the critical frequency of the ordinary ray usually denoted by a superscript *o* and the critical frequency of the extraordinary ray, denoted by the superscript *x*. Subscripts refer to the layer. For example, $f_{F_2^x}$ refers to the critical frequency of the extraordinary ray for the F_2 layer. For frequencies greater than 2.5 Mc the ordinary and extraordinary critical frequencies at Washington, D.C., are related approximately by

$$f^x = f^o + 0.8 \quad (12)$$

where the frequencies are in megacycles.

Normal ionosphere properties, *i.e.*, those whose variations may now be predicted with reasonable precision, have been emphasized above. In addition to these variations, there are those resulting from less easily predictable ionosphere disturbances: ionosphere storms, sudden ionosphere disturbances resulting in fade-outs and lengthy periods of absorption below the *E* layer.

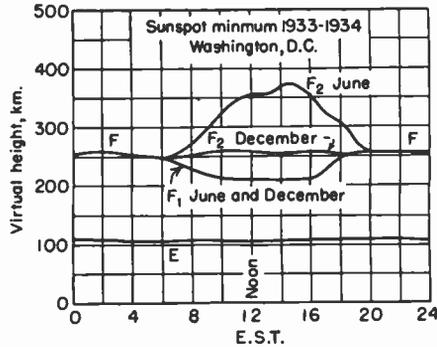


FIG. 24. Average diurnal variation in virtual heights of ionosphere layers during sunspot minimum. (*Bureau of Standards Ionosphere Reports, Monthly in Proc. IRE; Smith, Gilliland, and Kirby, Nat. Bur. Standards, J. Research, 21, 835, December, 1938; Gilliland, Kirby, Smith, and Reyrer, Proc. IRE, 26, 1347, November, 1938.*)

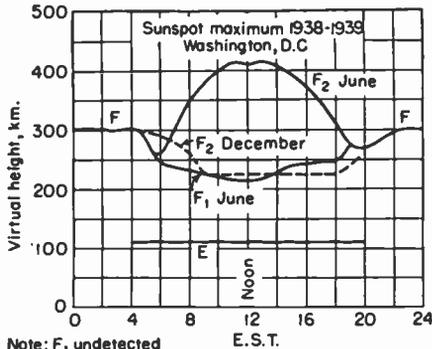
Considerable study is being devoted to such disturbances, but various investigators disagree as to the predictability of such effects.²

11. Noise. Among the factors limiting the usefulness of a received signal is noise, which may originate in any of the following places:

1. Within the receiver circuits
2. Within the transmitter circuits
3. Interfering signals
4. Atmospherics (static) and man-made noise

The principal concern here will be with item 4.

Noise Wave Forms. Noise, in the broadest sense, is any type of interference. It may include continuous signals from undesired transmitters, for which the noise is contained within a known, relatively narrow, band of frequencies, and also discontinuous noises, for which the frequency band occupied



Note: F_1 undetected in winter

FIG. 25. Average diurnal variation in virtual heights of ionosphere layers during sunspot maximum. (*Bureau of Standards Ionosphere Reports, Monthly in Proc. IRE; Smith, Gilliland, and Kirby, Nat. Bur. Standards, J. Research, 21, 835, December, 1938; Gilliland, Kirby, Smith, and Reyrer, Proc. IRE, 26, 1347, November, 1938.*)

is essentially infinite. Continuous disturbances are most easily studied by conven-

¹ Appleton, E. V., and G. Builder, Ionosphere as a Doubly-refracting Medium, *Proc. Phys. Soc.*, **45**, 208-220, Mar. 1, 1933.

² Kenrick, G. W., A. M. Braaten, and J. General, The Relation between Radio-transmission Path and Magnetic Storm Effects, *Proc. IRE*, **26**, 831-847, July, 1938; Dellinger, J. H., Sudden Disturbances of the Ionosphere, *Proc. IRE*, **25**, 1253-1290, October, 1937; Gilliland, T. R., S. S. Kirby, and N. Smith, Characteristics of the Ionosphere at Washington, D.C., June 1938, *Proc. IRE*, **26**, 1033, August, 1938.

tional methods. The term *noise* is commonly restricted to mean discontinuous disturbances.

Discontinuous noises may be considered to be made up of sharp pulses, the frequency with which the pulses occur determining the character of the noise. If the pulses are relatively infrequent and clearly separated, the noise is said to be *impulsive*. If, on the other hand, the pulses follow each other so rapidly that they overlap and are not clearly distinguishable, then the noise is *random*. Between these two types any gradations may occur. Ignition noise is impulsive. Tube and thermal agitation noises are random.

Since the frequency spectrums of discontinuous noises are infinite in extent, their magnitudes will depend upon the band width of the device with which they are measured. Jansky¹ and others have shown that the peak, average, and effective voltages of discontinuous noises depend upon band width in the manner shown in the table:

	Impulsive	Random
Peak.....	Proportional to band width	Proportional to $\sqrt{\text{band width}}$
Average.....	Independent to band width	Proportional to $\sqrt{\text{band width}}$
Effective.....	Proportional to $\sqrt{\text{band width}}$	Proportional to $\sqrt{\text{band width}}$

For thermal agitation noise, Jansky found the peak-to-effective voltage ratio to be 4 and the average-to-effective voltage to be 0.85.

Atmospheric noise resembles random noise in that the individual pulses overlap. While the measured voltages increase, therefore, directly as the square root of the band width, the ratios of peak-to-effective voltage are not constant as in the case of thermal noise.

The *maximum tolerable noise level* has not been measured under a wide enough variety of circumstances, up to the present time, to be able to specify its value for all conditions. The tolerable noise level depends upon a great many factors, including the following:

1. Type of service (sound, television, etc.)
2. Quality of service (excellent to poor)
3. Volume range of program material
4. Width of frequency band
5. Character of noise
6. Type of modulation (amplitude or frequency, preemphasized or flat, etc.)
7. Method of measurement

The effect of these factors has not been completely studied, particularly so far as correlating noise levels with the psychological irritation they produce to the ear (or to the eye in television and facsimile). The problem of adequate readability and articulation in the presence of high noise levels received much attention during the war, *e.g.*, at the Psychoacoustic Laboratory at Harvard University. The results of these careful investigations are not yet generally available, however.

12. Noise Measurements. Several commercial instruments of the semipeak vacuum-tube voltmeter type² are available for the purpose. The indicating instrument should have a natural period of 0.5 to 0.7 sec and a damping factor between 10 and 100 (ASA test methods). The actual time constants of the combination should be determined more by the electrical circuit than by the indicating instrument. The charging time of the circuit should be approximately 10 millisecc and the discharge time

¹ Jansky, K. G., An Experimental Investigation of the Characteristics of Certain Types of Noise, *Proc. IRE*, 27, 763-769, December, 1939.

² Burrill, C. M., Progress in the Development of Instruments for Measuring Radio Noise, *Proc. IRE*, 29, 433-442, August, 1941; An Evaluation of Radio-noise-meter Performance in Terms of Listening Experience, *Proc. IRE*, 30, 473-479, October, 1942.

approximately 600 millise^c.¹ The addition of a frequency weighting network to simulate the ear's response is sometimes recommended.

With an instrument of this type, the signal-to-noise ratios required at the output of a sound receiver for various qualities of service are, approximately,

Perfect signal.....	60-80 db
Good quality.....	40-50 db
Adequate intelligibility.....	15-30 db

For television, a peak-signal to peak-random noise ratio of 40 db gives a perfect picture, while a ratio of 30 db is intolerable. For single frequency noise, which is a small multiple of the line frequency, the interference is barely perceptible for a signal disturbance ratio of 50 db while for a ratio of 35 db it is intolerable.²

13. Atmospheric Noise. *Atmospherics*, or *static*, originates in lightning discharges. The impulses are frequent and overlap, so that the noise is more or less random, with

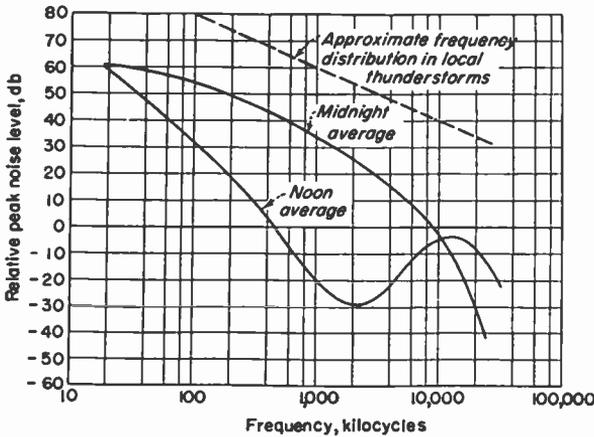


Fig. 26. Atmospheric static characteristics.

sharp peaks exceeding the average level. Atmospheric static originates both in local storms, relatively infrequent in northern latitudes, and in more distant tropical storm centers. Static is propagated in the same manner as other radio waves; variations in distant atmospherics may often be predicted on this basis.

The signal strength from local storms varies approximately inversely as the frequency.³ Thunderstorms and static are, of course, more intense in summer than in winter. The curves of Fig. 26 were measured near New York City, but probably are representative of most of the United States as well.

Since much static is of tropical origin, the lowest disturbance levels are found at distances remote from the equator, especially for low frequencies. However, at u.h.f. and at broadcast frequencies during the day, for which only short-distance communication is feasible, most static is of local origin. Since local storms are infrequent in northern latitudes—about 30 per year near New York City—static causes little interference with communication above 30 Mc. At 150 Mc the peak voltage from thunderstorms has been found to vary approximately inversely as the distance, being

¹ Aggers, C. V., D. E. Foster, and C. S. Young, Instruments and Methods of Measuring Radio Noise, *Trans. AIEE*, 59, 178-192, March, 1940.

² Jarvis, R. F. J., and E. C. H. Seaman, The Effect of Noise and Interfering Signals on Television Transmission, *POEE J.*, 32, 193-199, October, 1939.

³ Potter, R. K., An Estimate of the Frequency Distribution of Atmospheric Noise, *Proc. IRE*, 20, 1512-1518, September, 1932.

75 ± 10 db above 1 μv per m when measured with a 1.5-Mc band width at a distance 1 mile from a storm.¹

In the absence of either atmospheric or man-made static, Jansky has found that noise is still picked up by the receiver antenna, noise which he ascribes to stellar radiation. This noise, at frequencies between 9 and 21 Mc was found to be some 10 to 30 db above the level of thermal agitation noise in the receiver, and, except for man-made noise, is the limiting noise at high frequencies a large portion of the time.² Such "cosmic" noise is also believed to limit the sensitivity of v-h-f and u-h-f receivers.

14. Man-made Noise. This interference may be generated by internal-combustion-engine ignition systems, by power-line discharges, by diathermy machines, by motor-brush sparking, and by other electrical devices. Man-made noise usually reaches the receiver input in the following ways:³

1. By radiation or capacitance pickup by receiver antenna direct from noise source (relatively rare) or from power lines conveying voltages to vicinity of antenna
2. By transmission over power lines direct to receiver

Means for reducing such pickup are described in the literature.⁴

Typical noise voltages within the broadcast band measured at the terminals of noise-generating devices follow:⁵

Typical Noise Voltages

Source	Line-to-line r-f voltage, mv	Line-to-ground r-f voltage, mv
Vacuum cleaner.....	3	3.5
Electric razor.....	40	5.6
Diathermy machine.....	250	37
Portable electric tool.....	20	26

The band width of the measuring device was presumably between 3 and 5 kc. These voltages may be reduced by the addition of simple noise-suppression filters.

During the past few years interference of radiation from physician's diathermy machines with radio communication has become objectionable at high frequencies. Jansky⁶ has measured peak power levels of such interference ranging from 24 to 40 db below 1 μmW . The FCC is attempting to reduce this type of interference by assigning several frequency bands for the use of diathermy, r-f heating, and similar sources of interference. It is expected, however, that it will be some years before older machines not conforming to the regulations are taken from service.

At v.h.f. ignition and diathermy are the most objectionable types of noise. In cities diathermy signal strengths in excess of 100 μv per m are found. The peak ignition noise produced by 90 per cent of the vehicles passing 100 ft from an antenna 35 ft high has been found to be less than between 9 and 20 μv per m per kc l-f band width. The higher value is for 40 Mc, the lower for 450 Mc.⁷ Vertical polarization gives a little more ignition noise than horizontal. In New York City ignition-noise field intensities between 1 and 40 are found at typical antenna locations.

¹ Schaffer, J. P., and W. M. Goodall, Peak Field Strength of Atmospherics Due to Local Thunderstorms at 150 Megacycles, *Proc. IRE*, **27**, 202-207, March, 1939.

² Jansky, K. G., Minimum Noise Levels Obtained on Shortwave Receiving Systems, *Proc. IRE*, **25**, 1517-1530, December, 1937.

³ Block, L., Radio Interference, *Philips Tech. Rev.*, **3**, 235-240, August, 1938; Merriman, H. O., and F. G. Nixon, Radio Interference—Investigation, Suppression, and Control, *Proc. IRE*, **27**, 16-21, January, 1939.

⁴ For example, V. D. Landon and J. D. Reid, A New Antenna System for Noise Reduction, *Proc. IRE*, **27**, 188-191, March, 1939. A survey of noise-reducing systems is contained in the "Radio Noise Reduction Handbook," 1938, Radio, Ltd., Santa Barbara, Calif.

⁵ Haskins, R. L., and C. W. Metcalf, Station Coverage, *Communications*, **18**, 23-26, April, 1938; British Standard Specification for Limits of Radio Interference, *Brit. Standards, Spec.* 800-1937.

⁶ Jansky, *loc. cit.*

⁷ George, R. W., Field Strength of Motor Car Ignition between 40 and 450 Megacycles, *Proc. IRE*, **28**, 409-412, September, 1940.

CHAPTER 13

ELECTRON TUBES

By G. D. O'NEILL¹

INTRODUCTION

The scope of the present chapter is limited to devices in which conduction of electric current between two or more electrodes takes place as the result of the emission of electrons into space from one of the electrodes. Thus is ruled out consideration of such devices as those depending primarily upon the properties of semiconductors, despite the fact that such properties are of importance in electron tubes, especially with regard to the nature or quality of electron-emissive coatings.

There are several classes of devices which, on this basis alone, would require inclusion if further restrictions were not imposed. Among these would be certain cold-cathode, gas-filled tubes, thyratrons, pool-type rectifiers, and fluorescent lamps. Although some of these devices are of considerable interest in radio applications, discussion of their characteristics and design is omitted largely on the ground that information in greater detail than is given in the manufacturers' ratings is of value to a much smaller group of readers than is the case for high-vacuum electron tubes.

Except in a few instances, methods of testing for the various characteristics of electron tubes are not described, primarily for two reasons: there is now a plethora of texts in which the common tests are adequately described and, if the meaning of a characteristic is thoroughly understood—even if only from its definition—the engineer need not consult a book to find out how the test is to be made.

The extent to which various topics in this chapter are pursued depends largely upon relative need. By way of example, thermionic emission is the concern of more engineers than is photoelectric emission; more engineers are concerned with the processing, design, manufacture, and use of receiving tubes than of transmitting tubes.

Finally, the manner of treatment of the various topics is, it is hoped, such as to be of maximum benefit to the average engineer, visualized as having a fair background in physics and mathematics but without the advantages of graduate training or the equivalent in experience. Discussion of vacuum techniques and certain "practical" aspects of tube making such as glass sealing, exhausting, and coating techniques have not been included for the reason that no amount of reading on such topics could take the place of learning by even a small amount of experience in a laboratory or factory.

ELECTRON EMISSION

The passage of electrons from a solid into a gas or vacuum is called *electron emission*. The mechanism by which this passage or emission occurs in most cases classifies the general phenomenon into the categories thermionic, photoelectric, secondary, and field emission.

1. Thermionic Emission. The kinetic energies of the free electrons in a metal are distributed in accordance with the Fermi-Dirac relation shown in Fig. 1. Here the abscissas represent kinetic energy on a linear scale with zero energy at the origin. The ordinates may be taken as the probability of finding an electron with the corresponding energy.

Curve *a* shows the energy distribution of the free electrons at absolute zero. It is noted that the most probable energy corresponds to the voltage E_1 and that there are

¹ Research Department, Sylvania Electric Products, Inc.

no electrons having energies higher than this amount. At a higher temperature—say, 293°K or room temperature—the curve will look more like *b*; it is seen that a few electrons now have energies greater than E_1 . Curve *c* represents the energy distribution at a still higher temperature.

The energy E_2 of Fig. 1 represents a critical level. An electron arriving at the surface of the metal may emerge from it if the surface-directed component of its energy is equal to or greater than E_2 . If the surface-directed component is just equal to E_2 , the electron emerges with zero initial velocity; if it is greater than E_2 the excess is the measure of the initial velocity upon emergence.

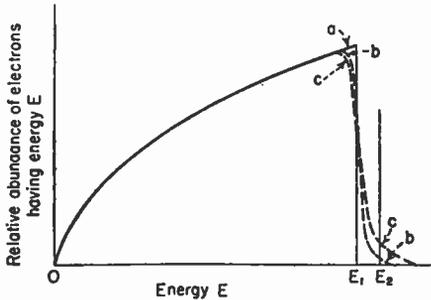


FIG. 1. Energy distribution of electrons in metals.

Not all electrons having a surface-directed energy greater than the work function are emitted; a fraction \bar{r} will be reflected from the surface.

The quantitative relation between these variables is expressed by the relation

$$I = A_0(1 - \bar{r})T^2e^{-b/T} \quad (1)$$

where I = emission current per unit area

A_0 = constant having a theoretical value of 120 amp/sq cm per deg² for metals

\bar{r} = reflection coefficient

T = absolute temperature

The exponent b in Eq. (1) is a measure of the work required to remove an electron from the cathode and is proportional to the work function of the cathode, the relation being

$$b = \phi e/k \quad (2)$$

where ϕ = work function

e = magnitude of charge of electron

k = Boltzmann's constant

Expressed in practical units,

$$b = 1.16 \times 10^4 \phi \quad \text{deg K} \quad (3)$$

For most practical cases, therefore, the thermionic-emission density for metals may be taken as

$$I = 60T^2e^{-11.600\phi/T} \quad \text{amp/sq cm} \quad (4)$$

when T is given in Kelvin or absolute degrees centigrade and ϕ is in volts.

Agreement between theory and practical results is usually satisfactory for metals, since a check on the value of ϕ from thermionic-emission data can be made by determining the long-wavelength limit of the photoelectric effect for the same cathode. With composite surfaces a number of factors commonly lead to inconsistencies, and recourse is generally had to statistical data where the ultimate objective is the design of an electron tube.

2. Tungsten Filaments. A brief summary of some practical thermionic cathode data is given below. These data are not, of course, intended to represent more than a small part of the information available in the complete literature.

Tungsten. Pure tungsten filaments were extensively used in early electron tubes. The ratio of emission to heating power is very low on account of the high work func-

tion of tungsten, consequently now only large transmitter tubes where conditions of temperature and voltage are severe employ tungsten.

The work function ϕ is given as 4.54 volts. Accordingly, the emission I_s in amperes per square centimeter may be calculated by Eq. (4) with the results tabulated for deg K, shown in the fourth column of Table 1.

Table 1. Properties of Tungsten and Thoriated Tungsten Wire

Temp. °K	ρ , ohm-cm $\times 10^3$	η , watts/cm ²	I_s (tungsten), amp/cm ²	I_s (thoriated tungsten), amp/cm ²	Expansion (L _T /L ₁₀₀₀ °K)	I' (I/d ^{3/2})	$E' \times 10^3$ (Ed ^{3/2} /L)
293	5.49	1.0000	0	0
500	10.56	9.71×10^{-3}	1.0009	47.62	0.640
1000	24.93	0.602	9.1×10^{-8}	1.0032	244.1	7.749
1100	27.94	1.03	2.5×10^{-8}	1.0037	301.0	10.71
1200	30.98	1.66	4.0×10^{-8}	1.0042	363.4	14.34
1300	34.08	2.57	4.0×10^{-8}	1.0048	430.9	18.70
1400	37.19	3.83	3.1×10^{-8}	1.0053	503.5	23.85
1500	40.36	5.52	9.4×10^{-8}	1.8×10^{-8}	1.0059	580.6	29.85
1600	43.55	7.74	9.27×10^{-7}	8.5×10^{-8}	1.0065	662.2	36.73
1700	46.78	10.62	7.08×10^{-6}	3.4×10^{-7}	1.0071	747.3	44.52
1800	50.05	14.19	4.47×10^{-5}	1.15	1.0078	836.0	53.28
1900	53.35	18.64	2.28×10^{-4}	3.55	1.0084	927.4	63.02
2000	56.67	24.04	1.00×10^{-3}	9.55	1.0090	1,022	73.75
2100	60.06	30.5	9.93×10^{-3}	23.44	1.0097	1,119	85.57
2500	73.91	69.8	0.298	1.0125	1,526	143.6
3000	95.76	185.8	14.15	1.0164	2,072	243.0
3655*	117.1	382.6	479.9	1.0221	2,838	423.4

* Melting point.

Thoriated Tungsten. The inclusion of a few parts per hundred of thorium oxide in tungsten wire permits, with proper processing, a considerable reduction in the work function. This processing consists in conversion of some of the thorium oxide to thorium metal which then diffuses outward, covering the tungsten. The emission from such a filament is given by Eq. (4) in which

$$b = 52,600 - 18,800\theta' \quad \text{deg} \quad (5)$$

where the factor θ' is related to θ , the fraction of the surface covered with a monatomic layer of thorium, as in Table 2; $A_0(1 - \bar{f})$ is taken as 60.2 amp/cm²/deg².

Table 2

θ	θ'	θ	θ'
0.00	0.00	0.5	0.790
0.05	0.127	0.6	0.863
0.1	0.240	0.7	0.920
0.2	0.430	0.8	0.960
0.3	0.579	0.9	0.988
0.4	0.697	1.0	1.000

It will be recognized that the temperature equivalent 52,600 in Eq. (5) is the value of b for pure tungsten. Therefore, for no thorium the work function is that of tungsten, 4.54 volts; for the fully activated surface, the work function will be about 2.9 volts.

3. Activation of Thoriated Tungsten Cathodes. Activation—the process of causing the thorium to diffuse out of the wire and coat the surface—is usually accomplished in one of two ways:

For *small filaments*, in which the operating current is of the order 0.25 amp or less, thorium oxide is reduced by raising the temperature of the filament to about 3000°K by the application of a voltage corresponding to two to three times that required for normal operation. After about 30 sec the voltage is reduced to about 1.5 times the normal value and held for about 10 to 15 min to permit diffusion of thorium to the surface of the wire. The word "about" has been used advisedly, since some experimentation is usually in order before a satisfactory "aging" schedule for a particular design of filament can be specified.

For *large filaments*, as in transmitting tubes, the thorium oxide is not sufficiently reduced by the heat-treatment described above; consequently resort is had to carburizing, in which the filament is glowed at high temperature in the presence of a hydrocarbon gas such as acetylene at very low pressure (or in a heavier compound diluted with hydrogen at atmospheric pressure). It is usually preferred that the pressure of the gas be restricted to an amount just sufficient to cause penetration of the carbon to the desired depth; a safer practice is to admit a still lower quantity and remove the resultant hydrogen before more acetylene is admitted. Carburization in the presence of an excessive amount of gas frequently results in uneven penetration and excessive brittleness. The acetylene must be highly purified and entirely free of oxygen or water vapor. Ordinarily, the carburization is carried to the depth representing about 20 per cent of the cross-section area. In this process the outer shell of the filament is converted to a very brittle carbide of tungsten having high electrical resistivity. During carburization the thorium oxide is reduced to metallic thorium which may be diffused to the surface by an aging schedule similar to that described for the small filaments.

4. Properties of Pure and Thoriated Tungsten Cathodes. Column 5 of Table 1 gives the emission of pure and thoriated tungsten. Pure tungsten filaments are ordinarily operated at temperatures of 2500 to 2570°K, thoriated tungsten at 1900 to 2000°K.

Table 1 provides data for the design of a tungsten filament operating in vacuum. Excluding second-order effects, such as temperature of the surroundings and the variation in temperature along the filament, the variables will be

L = length of filament	η = radiation intensity
E = filament voltage	r = radius of wire
I = filament current	d = diameter of wire
ρ = resistivity	

Formulas have been published for corrections due to cooling of the filament by the lead wires and hooks. Since the complete data necessary for precise determination of these corrections are seldom available, it is generally preferable, for ordinary computations, to follow a rule of thumb. This is simply to observe, in a sample or similar tube, the amount of filament which operates below visible incandescence and to consider that this part has zero resistance while the remainder has uniform resistance per unit length. The hot part of the filament is computed and the parts cooled by leads and hooks then added to give the final length. In the paragraphs to follow, L means the length of the hot part of the filament.

The hot resistance of the filament will be taken as the quotient of total filament voltage E by filament current I , and is related to the resistivity and dimensions by

$$E/I = \rho L/\pi r^2 \quad \text{ohms} \quad (6)$$

With the filament heated by power EI , the radiation intensity η , or power radiated per unit area, is

$$\eta = EI/2\pi rL \quad (7)$$

Since Table 1 gives ρ and η for centimeter units, L and r will be expressed in centimeters and the η will be in watts per square centimeter.

From Eqs. (6) and (7), the current and voltage of a wire 1 cm long and 1 cm in diameter may be computed for the temperatures given in the table, using the corre-

sponding values of ρ and η . These values, I' and E' , are given in the last two columns of the table. For values of I' and E' at temperatures intermediate between those given, interpolation between the next higher and lower values yields results which are sufficiently accurate for most purposes. Then, for a wire of any other diameter d and length L operating at the same temperature, the current I will be

$$I = I'd^{3/2} \quad \text{amp} \quad (8)$$

and E will be

$$E = E'L/d^{1/2} \quad \text{volts} \quad (9)$$

In practice, the temperature of a wire having known dimensions is determined by computation of I' from current and diameter or of E' from length, voltage, and diameter. The temperature is then found from the table. In design, the length, voltage, and temperature are likely to be specified, in which case the diameter will be obtained from Eq. (9) by substituting the value of E' corresponding to the desired temperature. This diameter is then substituted into Eq. (8) with the corresponding value of I' to obtain the required current.

The constants of Table 1 apply only to clean tungsten or thoriated tungsten, hence the values of η , E' , and I' do not apply to the heater wire of an indirectly heated cathode. However, Eq. (6) does apply in any case and is of use in computing the temperature of heater wire, when the design is achieved through cut-and-try methods. Here ρ is determined from voltage, current, and dimensions of the experimental tubes. The final tube should be so designed that the operating temperature of the heater does not exceed about 1450°K. Temperatures in excess of this value are apt to result in burnout trouble during exhaust or aging and may result in early failure of the heater.

It is difficult to measure with the degree of accuracy required the average diameter of a small tungsten wire when the wire is to be used in an electron tube or incandescent lamp, and recourse is had to the weight as measured by a sensitive torsion balance. Except for large sizes, the wire is commonly purchased and specified in terms of its weight in milligrams for a 200-mm length, the tolerance being specified—commonly 2 per cent. Table 3 gives the data required for the conversion of diameter in millimeters and mils (thousandths of an inch) to weight in milligrams per 200 mm.

5. Carburized Filament Design. In designing a carburized filament, it may be considered that the electron emissivity is the same as for the uncarburized thoriated wire. The resistance depends on the depth of carburization. As an approximation, the carburized portion may be taken to have infinite resistivity; *i.e.*, in the computations the resistance is taken on the basis of the uncarburized core while the radiation is taken for the original diameter. At the normal operating temperature of 1900 to 2000°K, the radiation intensity of the carburized wire is about 20 per cent greater than for clean wire at the same temperature. Thus, when the wire is 20 per cent carburized, a given current heats both to approximately the same temperature, the voltage being about 20 per cent greater for the carburized filament.

6. Oxide-coated Cathodes. The oxide-coated or Wehnelt cathode comprises a base metal such as nickel, tungsten, or one of a large number of alloys, coated with the oxides of barium and strontium or of barium, strontium, and calcium. There are numerous other types of coated cathodes, but their commercial importance is small relative to that of the BaO-SrO variety at the time this is written.

Because of the wide variety of test results obtained when attempts are made to evaluate A_0 and b of Eq. (1) by means of experimental data of emission vs. temperature, it is safer to evaluate coated cathodes by performance records at a specified value of heating power per unit area than from the constants of the emission formula. The work function of thermionic cathodes may, of course, be determined from measurements of the long-wavelength limit of the photoelectric effect, but such results do not usually correlate well with thermionic data.

In the case of receiving tubes employing indirectly heated cathodes, the ratio of heater power in watts to area in square centimeters is usually about 2.75 to 3.0; in

high-vacuum rectifiers it is higher, sometimes as great as 4.0 watts per sq cm to provide for the high peak currents frequently required.

Table 3. Weight and Diameter of Tungsten Wire

Weight, mg/200 mm	Diameter, mm	Diameter, mils	Weight, mg/200 mm	Diameter, mm	Diameter, mils
1	0.01830	0.72	32	0.1035	4.08
1.5	0.02242	0.88	33	0.1050	4.14
2	0.02589	1.02	34	0.1066	4.20
2.5	0.02894	1.14	35	0.1083	4.26
3	0.03167	1.25	36	0.1097	4.32
3.5	0.03425	1.35	37	0.1113	4.38
4	0.0366	1.44	38	0.1127	4.44
4.5	0.0388	1.53	39	0.1142	4.50
5	0.0409	1.61	40	0.1157	4.56
5.5	0.0429	1.69	41	0.1172	4.61
6	0.0448	1.765	42	0.1185	4.67
6.5	0.0467	1.84	43	0.1199	4.72
7	0.0484	1.905	44	0.1213	4.76
7.5	0.0501	1.97	45	0.1227	4.83
8	0.0517	2.04	46	0.1240	4.89
8.5	0.0533	2.10	47	0.1254	4.94
9	0.0549	2.16	48	0.1267	4.99
9.5	0.0564	2.22	49	0.1280	5.04
10	0.0578	2.28	50	0.1294	5.09
11	0.0607	2.39	55	0.1357	5.34
12	0.0634	2.495	60	0.1417	5.58
13	0.0659	2.60	65	0.1474	5.81
14	0.0684	2.69	70	0.1531	6.03
15	0.0708	2.79	75	0.1584	6.23
16	0.0732	2.88	80	0.1536	6.44
17	0.0754	2.97	85	0.1687	6.64
18	0.0776	3.055	90	0.1735	6.83
19	0.0798	3.14	95	0.1783	7.02
20	0.0818	3.22	100	0.1830	7.20
21	0.0838	3.30	110	0.1919	7.56
22	0.0858	3.38	120	0.2005	7.89
23	0.0878	3.455	130	0.2087	8.22
24	0.0897	3.53	140	0.2165	8.53
25	0.0915	3.60	150	0.2242	8.83
26	0.0935	3.67	160	0.2314	9.11
27	0.0951	3.75	170	0.2387	9.39
28	0.0968	3.81	180	0.2455	9.67
29	0.0985	3.88	190	0.2523	9.93
30	0.1002	3.95	200	0.2589	10.18
31	0.1018	4.01			

7. Emission Efficiency. It is difficult to decide what one would call an "average" oxide-coated cathode. On the basis of usual experience, a fair comparison might be made of the emission efficiencies—the quotient of emission current in amperes by heating power in watts—between the three principal types of thermionic cathodes. The quantities given are taken at the ordinary level of emission density of an oxide-coated receiving tube cathode.

Emitter	I_a /sq cm	Watts/sq cm	I_a /watt
Oxide-coated.....	0.10	2.75	0.0364
Thoriated tungsten.....	0.10	8.0	0.0125
Tungsten.....	0.10	55	0.0018

In designing an indirectly heated cathode for general-purpose and receiving tubes, it is generally desirable to make the ratio of length to diameter as great as possible, since end-loss is then low and clearance between cathode and the next nearest electrode

is maximum. On the other hand, provision must be made for insertion of the heater, and the length-to-diameter ratio must not be so great that bending or bowing of the cathode is likely to be serious. In practice, the ratio is generally about 25.

8. Flat or Ribbon Filaments. Where a cathode with a flat or nearly flat surface is required, such as in a cathode-ray tube or others in which the electron stream is formed into a cylindrical beam, the heater-power efficiency is low compared with that of a cylinder coated on the side, although this is frequently of minor interest. Where a new design is required, life tests are commonly made over a range of heater power, and the heater current rating is selected on the basis of life test results.

For the directly heated cathode, the design of a filament is made on the basis of performance records of tubes using the same material. Knowing the dimensions of the filaments of these tubes and the voltage and current for satisfactory life, the values of ρ and η are computed. The new filament, if round, may be computed as for the tungsten wire; corrections for the cooling effect of the leads and hooks are applied in the same manner.

The use of a ribbon filament provides an additional degree of freedom in choice of dimensions or voltage or current, since the surface area may be varied without changing the area of cross section.

The relations corresponding to Eqs. (6) and (7) for the ribbon filament will be

$$\rho = abE/IL \quad \text{ohms} \tag{10}$$

$$\eta = EI/2(a + b)L \quad \text{watts/sq cm} \tag{11}$$

where a and b are the width and thickness of the ribbon.

From the data on known or experimental tubes, the values of ρ and η are computed for the conditions that have been shown to give best results for life test. Then when values of length, current, and voltage are specified, the other dimensions will be

$$a = \frac{EI}{4\eta L} + \frac{1}{2} \sqrt{\left(\frac{EI}{2\eta L}\right)^2 - \frac{4\rho LI}{E}} \quad \text{cm} \tag{12}$$

$$b = \frac{EI}{4\eta L} - \frac{1}{2} \sqrt{\left(\frac{EI}{2\eta L}\right)^2 - \frac{4\rho LI}{E}} \quad \text{cm} \tag{13}$$

Since the thickness b is important and is the difference between two quantities apt to be nearly alike, computation should be carried out by the use of log tables in which the values of the logarithms are given to at least six digits. The result is conveniently checked by substitution in Eq. (9). A negative or complex solution indicates the values chosen for E or I are too small, or that L is too great. In the practical case, the ratio a/b is desirably greater than 3 and less than 15.

Since ribbon filament is ordinarily made by rolling a round wire, the thickness is apt to be somewhat variable across the width. For this reason, ribbon filaments are ordinarily specified in terms of width (mils) and weight (mg/200 mm) to ensure against wide variations in current and temperature.

9. Preparation of Cathodes. The basic material, whether it is wire, ribbon, or an indirectly heated cathode, is thoroughly cleaned by solvent materials such as dichloroethylene, then heated in a hydrogen atmosphere to a temperature of about 1000°C before coating. The coating is usually applied in one of three ways:

Bench coating consists in passing the wire or ribbon alternately over revolving wheels in a cup of coating, then through an electric furnace. The coating is thus built up in several layers. The carbonates are suspended in a solution of barium nitrate, the nitrate serving as a temporary binder. The furnace is supplied with a stream of carbon dioxide.

Spray coating consists in spraying the coating suspension on the ribbon or cathode with a spray gun, nitrocellulose being used as a binder. The desired texture of the coating is achieved through adjustment of nozzle, air pressure, and distance between nozzle and cathode.

Electrophoresis consists in applying the coating to the cathode by passing the wire

through, or dipping the cathode into, a suspension of the coating and applying a strong electric field between cathode and metallic container.

In coating the cathode, the active materials are in the form of carbonates which are broken down to the oxides during the exhaust process. These carbonates, usually of barium and strontium, are ordinarily prepared especially for the purpose. Many of the larger tube-manufacturing firms prepare their own carbonates, while others purchase them from chemical supply houses in a form especially adapted to this work. The barium and strontium carbonates—sometimes calcium carbonate is also a constituent—are precipitated together from a hot solution of the nitrate by the addition of a solution of ammonium carbonate, sodium carbonate, or sodium bicarbonate, then washed and calcined. The temperature and concentration of the solution are closely controlled to produce crystals of the proper form; these crystals ordinarily have a diameter of a few microns.

10. Coating Suspensions. The composition of the suspension will depend upon the method of coating, the size of the wire, and the intended use of the cathode; the experience of the process engineer is also an influencing factor. For this reason, the coating formulas given below merely illustrate the way in which such suspensions are compounded; they are, however, practical for use in making experimental tubes.

1. Bench-coating formula:
 - 1,000 g carbonate
 - 1,000 cc distilled water
 - 100 g barium nitrate (decrease for wide ribbon)
2. For use in spray gun:
 - 575 g carbonate
 - 1,000 cc amylacetate
 - 20 g nitrocotton
3. Electrophoretic suspension:
 - 50 g carbonate
 - 5 g barium nitrate
 - 1,000 cc amyl acetate
 - 63 g nitrocotton

In all cases the suspension is ball-milled for a minimum of 12 hr to ensure complete separation and wetting of the individual particles. In using formula 1, it is necessary to heat the wire to a temperature high enough to fuse the nitrate or to convert it to the carbonate. In the other cases, care is taken to ensure thorough evaporation of the amylacetate or other solvents. Plasticizers are frequently added in small proportions to give a degree of flexibility to the coating when nitrocotton is used as the binder.

11. Exhaust and Aging of Coated Cathodes. During the exhausting of the tube, the carbonates are broken down to the oxides by the application of heat, the reaction being typified by the chemical equation



At atmospheric pressure the reaction does not proceed very far due to equilibrium between the carbonate, oxide, and carbon dioxide; at exhaust the carbon dioxide is continuously removed, permitting the reaction to go to complete breakdown of the carbonate.

After the tube has been sealed off, it is ordinarily aged. The cathode is heated by the application of a heater voltage higher than normal. This may be as great as two to three times normal operating value, followed by a second step in which the heater voltage is held at about 150 per cent of normal with the cathode passing a thermionic current of about 100 ma per sq cm. The emission current is usually limited by an incandescent lamp ballast in the plate circuit. The time required for aging may be a few minutes for the first step and 10 to 20 min for the second step. Aging schedules very widely depending upon such factors as the type of tube, the exhaust schedule, coating texture, and the composition of the base material. Their specification is ordinarily the result of cut-and-try experimentation.

12. Base Materials. For either filamentary or indirectly heated cathodes the base materials are rarely pure metals, notable exceptions being the tungsten wire used in some of the low-drain, subminiature types and the electrolytic nickel indirectly heated cathode often used in tubes wherein the production of free barium must be kept low to reduce the likelihood of grid emission.

The impurities added to the nickel base, or left there during refining, serve two purposes. In the first place, the cathode is assumed to have on or in the coating a small quantity of free barium which may be produced by the reduction of barium oxide through interaction with another metal. That nickel serves here to a certain extent may be concluded from the fact that the filament of 95 per cent Pt, 5 per cent Ni used in long-life repeater tubes has better emission characteristics than a pure platinum filament. However, the inclusion of other and more active materials permits the tube to be aged to a higher degree of activity, and in less time, than in the case where pure nickel is used. Such materials include magnesium, aluminum, silicon, titanium, and others, either alone or in combination.

Another reason for the inclusion of impurities, when the cathode is a nickel filament or ribbon, is to reduce *creep* during life. The impurities present at the grain boundaries of the metal prevent or reduce the slipping at the grain surfaces that usually occurs in a hot wire of pure metal, owing to the tension of the supporting hook. Thorium oxide was used in tungsten lamp filaments to prevent creep or sag before it was discovered that thorium could be diffused to the surface with resulting increase in electron emission.

Inclusions that are best for reduction of creep are not always best for aiding in the activation of the oxide coating. Aluminum serves both purposes rather well, but aging is often found to be critical. Magnesium also is effective and is less critical than aluminum, although it has been noted that it may become exhausted at an earlier period in life with accompanying reduction in the emission of the cathode.

The nature of the activator and the manner in which the breakdown of the carbonates is accomplished, *i.e.*, the temperature of the cathode and the amount and kind of gas present at the time, are factors which determine the nature of the interface between coating and base. A poor interface will be one having high resistance; where high peak current densities are required, as in pulsed operation, the voltage drop across the interface may be a considerable fraction of the anode voltage. Specialists in this field do not appear to be able to agree upon any one best activator. Practical or economic aspects may result in a conclusion different from what might be reached when long life is the primary objective. Considering such factors as the heat of formation of the oxide, rate of diffusion out of the base, and rate of diffusion through the BaO-SrO coating, the use of magnesium has been recommended for commercial receiving tubes or others where a high level of emission is important. Where long life is a criterion and filament heating power is of less importance, the 95 per cent Pt, 5 per cent Ni is still acceptable. The use of tungsten as a base metal for oxide-coated filamentary cathodes has increased notably in recent years. Most of the tubes used during the war in proximity fuses had such filaments, and these have been adapted to peacetime use in the subminiature tubes. A useful life of several thousands of hours has been reported.

For tubes designed for operation under especially rigorous conditions, modifications of the ordinary practices outlined above are required. For example, in many of the high-powered cavity magnetrons, in which emission is drawn from the cathode in pulses of one or a few microseconds duration, the required emission may be of the order of 50 or 100 amp per sq cm without resort to the use of abnormally high temperatures. Such cathodes are commonly made by first wrapping the nickel sleeve with nickel mesh made of fine wire; the mesh is then sintered to the sleeve. The coating may consist of the ordinary mixture of barium, strontium, and calcium carbonates with the addition of finely divided nickel powder in amounts up to as much as 30 per cent by weight. The coating is applied with a spatula or paddle. The nickel mesh is helpful in reducing the resistance of the interface or thin film of slags, such as barium silicate, which form over the base metal; the nickel powder is included primarily to

reduce sparking between cathode and anode. Another form of thermionic cathode for operation under severe conditions is the sintered thoria cathode, made by fusing thorium oxide with a refractory metal such as molybdenum, the latter in amounts up to about 30 per cent by weight.

13. Testing Thermionic Emission. The thermionic-emission properties of a tube are measured in a number of different ways, the choice being determined by the type of tube or the purpose of the test. A number of these tests will be described briefly. In most cases all electrodes except cathode are tied together at the socket to form a composite anode, and the anode current is noted under conditions which may be specified.

Ideally, the (composite) anode current of a tube should increase with the three-halves power of the anode voltage, when the cathode temperature is constant, until temperature-limited emission is reached.¹ From this point the current should increase only gradually as the field at the cathode is increased. In the actual case, especially with coated cathodes, it is often difficult or impossible to ascertain the point at which the transition from space-charge-limited to temperature-limited current occurs because of lack of uniformity in the temperature and emission constants of the cathode. Results are more meaningful if plotted as a Schottky line from which the true value of emission may usually be inferred.

The Schottky relation may be written in the form

$$I = I_s e^{\frac{e\sqrt{-F}}{KT}}$$

where I = anode current

I_s = emission at $F = 0$

T = absolute temperature of cathode

F = electric field at cathode surface

$e/k = 11,606$ (degrees per volt)

In order to use the Schottky equation for determining emission, it is not necessary to know the temperature of the cathode or the intensity of the field. Taking the logarithm of both sides of the equation,

$$\log I = \log I_s + \frac{5046}{T} \sqrt{-F}$$

The field F at the cathode is not easily calculated, but when the voltage is well above that required to produce temperature saturation, the field may be taken as proportional to minus the anode voltage. Thus the relation may be written

$$\log I = \log I_s + k \sqrt{e_p}$$

in which k is constant for a tube at constant cathode temperature and e_p is the plate voltage. Two things are now evident:

1. If $\log I$ is plotted against $\sqrt{e_p}$, a straight line is obtained at values of e_p greater than required for temperature saturation.

2. If the straight line is extrapolated to $e_p = 0$, the intercept on the vertical axis gives the logarithm of the emission.

For tubes having coated cathodes, care must be taken to avoid overheating. In many cases the emission may fall because of the decomposition of impurities on the anode so that materials poisonous to the cathode are released. For these reasons, approximation methods are frequently employed.

Some direct emission tests are as follows:

1. For small tubes having tungsten filaments, the emission current is measured at rated filament voltage and with the composite anode voltage at or near the value required to produce temperature-limited emission. Care must generally be exercised in keeping the time of test to a minimum to avoid overheating the electrodes.

2. For small tubes having coated filaments or coated, indirectly heated cathodes, the composite anode voltage is commonly set at a value which, in a well-activated tube, produces an emission current of approximately 200 ma per sq cm of emitter surface. The test must not be too long continued, although, for purposes of quality

¹ See discussion of *perveance* (p. 563) in which it is shown that the space-charge-limited current varies as the three-halves power of the voltage between anode and *outside* of the cathode.

control, it is made for a long enough time—say 3 to 10 sec—to ensure that the emission is reasonably stable.

3. Where a very high degree of cathode activity is required and where the application of a voltage sufficiently high to result in a current near the temperature-limited value would result in serious injury to the tube, the anode voltage is pulsed at a relatively low duty cycle.

4. For many purposes, the “flection-point” emission is taken as a criterion of cathode activity. The flection point is defined as the point on the diode characteristic (current plotted against composite anode voltage) at which the second derivative of current with respect to voltage reaches its maximum negative value. This point is often referred to as the “knee of the curve.” Although not rigorous, it is conveniently taken as point *x*, Fig. 2, the point of tangency between the diode characteristic and a straight line through the origin. Data may be taken point by point or by means of a cathode-ray oscillograph.

Indirect checks depend for their value upon the increase in emission with temperature, the total emission or tube performance being measured at a cathode temperature lower than normal. Two such checks will be described.

1. In tubes having filamentary cathodes that might be injured by the passage of a relatively high emission current, the check consists in measuring the filament voltage required for a substantially temperature-limited emission current of a specified magnitude. For such a test the current will ordinarily be approximately that of the total cathode current under ordinary operating conditions.

2. For large transmitting tubes, an indirect emission check consists in testing the power output of the tube operating as a self-excited oscillator with standard voltages, then reducing the filament voltage until the power output is decreased to a specified fraction of the original value. The filament voltage reading under this condition then constitutes an indirect check upon emission. The value of such a check depends, to a considerable degree, upon judicious choice of circuit conditions and is applicable only in comparing individual tubes of the same type.

14. Photoelectric Emission. The energies of the conduction electrons of a solid material may be raised to a magnitude sufficiently high to permit their emergence from the body by means of radiation without heating. When emission occurs through the agency of radiation, usually within or near visible wavelengths, the emission is classified as *photoelectric emission*, the emitter is a *photocathode*, and the electron tube, whether gas-filled or vacuum, employing a photocathode is a *phototube*.

For what one might loosely call “white light,” the magnitude of the photoelectric current for unit emitting area is proportional to the intensity of the light falling upon the photocathode and depends upon the work function of the surface in much the same manner as in the case of thermionic emission. On the other hand, the magnitude of the current also depends upon the color composition of the incident light for the

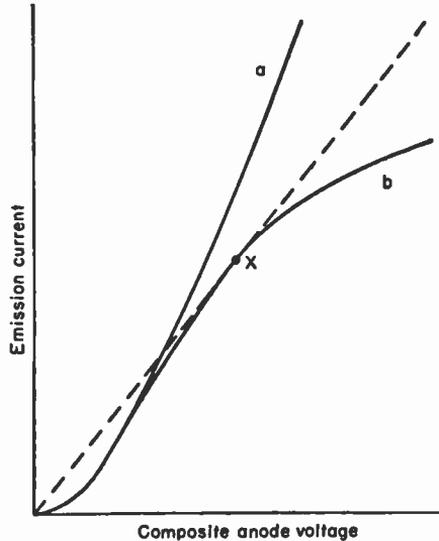


FIG. 2. Flection-point emission. Line *a* is hypothetical curve of space-charge current plotted against composite anode voltage. In the actual case the current is limited by emission line *b*. Tangent point *x* is taken as flection-point emission.

reason that the energy which the light may give to an electron is inversely proportional to its wavelength. For a given emitter there will be a maximum wavelength, called the *threshold wavelength*, at which photoelectric emission is obtained; for shorter wavelengths the excess energy imparted to the escaping electron beyond that required to overcome the surface forces will be the kinetic energy the electron will have upon emergence. Quantitatively stated,

$$hc/\lambda = \phi e + mv_0^2/2 = h\nu \quad (14)$$

where h = Planck's constant

c = velocity of light

λ = wavelength of radiation

ϕ = work function of photocathode

e = electron charge

m = mass of electron

v_0 = velocity of electron upon emergence¹

ν = wave frequency = c/λ in air or vacuum

The threshold wavelength is, therefore, the wavelength at which the second term of Eq. (14) just becomes zero. When the constants are evaluated in practical units and the work function ϕ is expressed in volts, the threshold wavelength λ_0 is given by

$$\lambda_0 = 12,395/\phi \quad \text{A} \quad (15)$$

where the wavelength is measured in angstrom units, $\text{A} = 10^{-8}$ cm.

As an example, if a photoelectric current is to be obtained which is sensitive to all the visible colors, and the limit of visibility is taken as 7000 A at the red end of the spectrum, the work function of the photoelectric cathode must be not greater than $12,395/7,000 = 1.77$ volts. Many phototubes have their long-wave limit extending well into the infrared region. These tubes are not, however, to be confused with the infrared cells such as the thallous sulfide type which are sensitive at wavelengths from within the visible range to and beyond 12,000 A.

While Eqs. (14) and (15) are intended only to be rigorous at absolute zero, the result is not greatly different at room temperature for the reason that the chance that a given electron shall be excited by the radiant energy appears to be independent of the energy it already had. Reference to Fig. 1 shows that the average energy is not greatly altered as the temperature is raised. One may concede that a few electrons may be emitted at higher temperatures for a wavelength slightly greater than that given by Eq. (15), but these electrons are more likely to constitute thermionic emission.

Most photoelectric cathodes exhibit a spectral selectivity such that the maximum response at a given level of incident radiant energy is obtained at a wavelength of or fairly near $2\lambda_0/3$. In the case of thin composite surfaces, such as those comprising an oxidized silver base upon which an alkali metal such as cesium or rubidium is deposited, two and sometimes three maxima are found, each being characteristic of an element or compound subjected to the radiation.

The "complete" photoelectric emission is expressed by an equation identical in form with that for thermionic emission. If the photoelectric cathode is exposed to the total thermal radiation from a black body at temperature T , the total photoelectric current for clean metals varies with T in accordance with the relation

$$I = A_1 T^2 e^{-h\nu_0/kT} = A_1 T^2 e^{-b_0/T} \quad \text{amp/sq cm} \quad (16)$$

in which A_1 is found to have values usually of about 10^{-10} to 10^{-9} amp per sq cm per deg² and $\nu_0 = c/\lambda_0$. The numerical value of b_0 is $1.16 \times 10^4 \phi$ deg as in the case for thermionic emission. This could be expected since the work ϕe done by the electron in leaving the cathode is the same regardless of whether its energy within the body is of thermal or photonic origin.

¹ The velocity v_0 will be greater by a small amount when the energy before excitation is added to both sides of the equation.

15. Photocathodes. The choice of a photoelectric tube depends, naturally, upon the use for which it is intended. The best guides in the choice of tube are the catalogues or handbooks of the most reliable manufacturers of such devices. The brief description to be given of some kinds of phototube cathodes is intended only for the general reader.

Of rather specialized interest are solid metal cathodes or cathodes coated to a relatively great thickness—*e.g.*, 100 or more atomic diameters—with metals having relatively high work function, such as zirconium, tantalum, or thorium. These are of interest in the measurement of ultraviolet light and usually are mounted in bulbs transparent to the desired regions of ultraviolet or employ quartz windows. Chemical cleanliness of the materials is of greatest importance; proper bulb baking and outgassing by induction heating are required for reasonably reliable results.

The sensitivity of some of these surfaces may be considerably enhanced and the long-wave limit λ_0 raised by passing a glow discharge between cathode and anode in the presence of hydrogen. This is preferably done after outgassing, the hydrogen being then pumped out before the tube is sealed off. Gases other than hydrogen, *e.g.*, sulfur vapor, have been tried with success.

Of greater interest are the photocathodes produced by the deposition of thin films of certain alkali metals upon a solid metal support. Because of the low work function of these surfaces, some of them are sensitive in the visible regions of the spectrum. A summary of features of present interest concerning some of these surfaces is shown below, the values given being approximate.

Surface	λ_m , A	λ_0 , A
S-1 (commonly Co-O-Ag).....	8,000	12,000
S-3 (commonly Rb-O-Ag).....	4,300	9,000
S-4 (commonly Cs-Sb).....	4,000	7,000
S-5 (commonly Cs-Sb)*.....	3,400	7,000
S-8 (commonly Cs-Bi).....	4,200	8,000

* In ultraviolet transmission bulb.

λ_m is wavelength for maximum sensitivity; λ_0 is long-wavelength limit.

Among the most useful commercial photoelectric cathodes, from the standpoint of sensitivity, is the Cs-CsO-Ag emitter. Although the maximum sensitivity of this surface does not correspond to that of the human eye—none of them do—its performance under tungsten or daylight illumination is of considerable use in commercial devices. It consists essentially of a well-cleaned silver support which, during exhaust, is outgassed, then oxidized by the passage of a glow discharge in oxygen. Care must be exercised in obtaining the proper depth of oxidation, the progress being observed by watching the subtractive interference color changes. The uncombined oxygen is then removed and cesium distilled into the bulb, care being taken to admit only as much cesium as is required to produce the proper sensitivity.

The response characteristics of some commercial phototube surfaces are shown in Fig. 3, wherein the average spectral response is compared with that of the average human eye. Relative sensitivities for each curve are plotted against wavelength for equal values of radiant flux at all wavelengths at the photosurface except for the S-5 photocathode, which is plotted for constant flux outside the bulb.

The photoelectric mosaics used in television camera tubes generally have the Cs-CsO-Ag emitting surface, the silver being in the form of tiny globules dispersed over the surface of a thin, nonconducting support, such as mica, through which capacitive coupling is obtained to a solid metal backing. During exhaust, the surfaces of the globules are oxidized and sensitized as described in the preceding paragraph.

16. Multiplier Phototubes. A multiplier phototube comprises essentially a photocathode, an anode, and one or more dynodes¹ sealed in an evacuated transparent

¹ Secondary electron-emitting electrode.

envelope or an envelope having a transparent window. Commercial multiplier phototubes commonly have a sensitivity of 5 to 10 μa per microlumen. Cesium-bismuth and cesium-antimony photocathodes are commonly used.

In operation, the first dynode is maintained at a potential positive with respect to the photocathode, and each succeeding dynode is positive with respect to that preceding. Electrons photoelectrically emitted from the cathode are drawn to the first dynode; with a secondary emission ratio greater than unity, secondary electrons are emitted which, in turn, give rise to an amplified current from the next dynode. By the time the electron stream has reached the anode, the original photocurrent is amplified δ^n -fold, where δ is the secondary electron emission ratio of the dynode and n is number of dynodes.

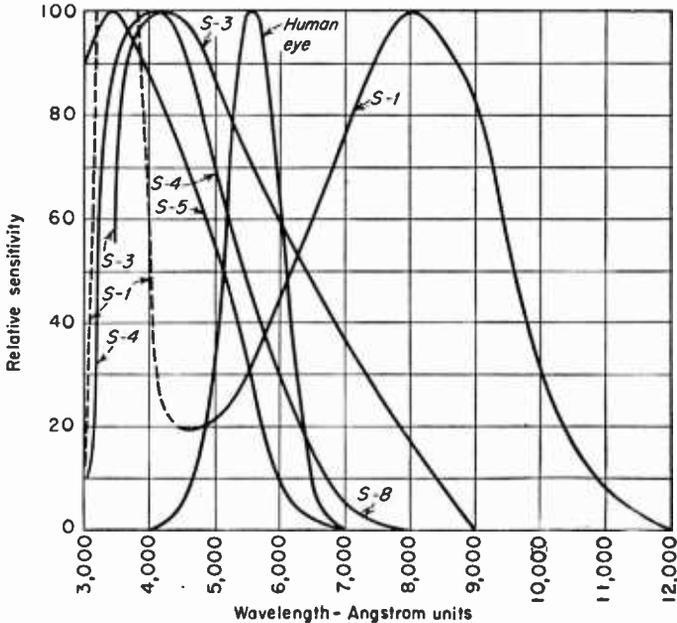


FIG. 3. Average response of commercial photocathode surfaces compared with that of the average human eye.

The multiplier phototube has a number of advantages compared with a phototube and separate amplifier. Since the amplification is entirely that of a single primary stream of electrons, the incident light which actuates the current may be relatively constant or it may be chopped, as by means of a rotating shutter, and the gain remains constant. Of still greater advantage, especially for the measurement of light at low intensity, is the high signal-to-noise ratio obtained.

In considering the signal-to-noise ratio of the multiplier phototube, it should be noted that total absence of light does not result in zero anode current. The small current observed under such conditions is known as the "dark current," the principal components of which are (1) amplified thermionic emission, (2) leakage current, (3) regenerative ionization. At high levels of cathode illumination, the dark current is likely to be of minor importance; at low levels, such as in astronomical work, the dark current may be important, although the resultant "noise" is still low compared to what would be obtained in a unit containing a phototube and separate amplifier having the same gain and a band width equal to that of the equipment into which

the multiplier phototube operates. At a very low level of illumination the photoelectric cell with separate amplifier becomes useless; thermionic and leakage currents, the Johnson noise of the coupling resistor, and the shot noise of the first amplifier tubes would prevent detection of the amplified photoelectric current. In the multiplier phototube, noises from such causes are excluded and the dark current may be reduced to extremely low values by cooling the tube and using conservative electrode voltages.¹ Engstrom² has discussed the manner in which the dark current for a cooled tube may be calibrated in such a manner that the lower limit of detectability may be reduced through a pulse-counting technique to about 2×10^{-16} lumen, the equivalent of photocathode current of 2×10^{-21} amp.

17. **Secondary Emission.** A third way in which electron emission may be produced is by the bombardment of a conducting or semiconducting solid with electrons or positive ions, known as *secondary emission*. Such an emitter, when used as an electrode of an electron tube, is called a *dynode*.

Secondary emission may be obtained by electron bombardment of pure metals, although the yield is not great. In this case a part of the energy of the bombarding or *primary electron* is transferred to one or several conduction electrons of the solid. If the surface-directed energies of the newly excited electrons are greater than ϕ_e , the energy required for their exit, they may be emitted as *secondary electrons*. In the case of pure metals, the ratio of secondary electron current to primary electron current, δ , is not appreciably affected by temperature.

When the secondary-electron-emitting electrode or dynode is a composite surface properly prepared for use as such, δ may be made considerably greater than unity with the result that a weak electron current, such as that from a photocathode, may be amplified many times by secondary emission. Other uses of the phenomenon are found in the dynatron and in the orbital-beam tube. The secondary emission ratio of such dynodes is not entirely independent of temperature.

The mechanism of secondary emission from composite surfaces is generally considered to be different from that of pure metals. In the latter case the "free" electrons of the metal are thought to provide the secondary electron current. In the case of composite surfaces, certain compounds of the metals, especially those of the second and third columns of the periodic table, provide much higher values of δ than do the pure metals. Among the best compounds are barium oxide and magnesium oxide; sulfides, halides, etc., of these metals are not satisfactory. Here the secondary emission is provided by the excitation, due to the energy of the incident primary electron, of the electrons of the outermost shells of atoms constituting the emitting material. In such a case it is considered that the permitted energy bands for electrons in the compound are either entirely empty or completely filled, and that a bombarding primary electron can lose its energy by causing the transition of an electron from a filled band to an empty band. The chance that an excited electron shall leave the body depends upon the position of the empty band with respect to the occupied band and the potential outside the body. If the empty band falls above the potential barrier going from the body to vacuum in the energy scheme—*i.e.*, if the transition to the higher band involves an amount of energy greater than ϕ_e —the excited electron may be emitted as a secondary electron. For an empty band below the barrier, the electron may fall back to its original energy level or it may be excited by another primary electron to a still higher band with the possibility of emission. Since the likelihood of this second excitation is not great, the relative difference in the value of δ for different substances appears to depend largely upon the relation between the lowest unoccupied energy band of the molecule and the work function of the surface.

If all of the energy brought into the body by the bombarding primary could be transformed into the energy required for the emission of secondary electrons, the ratio

¹ By using solid CO₂ as the coolant, the escaping gas may be used to prevent the condensation of moisture on the bulb.

² Engstrom, R. W., Multiplier Phototube Characteristics: Application to Low Levels, *J. Optical Soc. Am.*, 37 (6), 420-431, June, 1947.

δ would be very high. If V is the voltage between dynode and cathode, the energy of the primary just inside the dynode would be $Ve + \phi_e$; the energy would be transferred to n secondary electrons having a total energy $n\phi_e$. Thus the maximum number n that could be emitted would be $n = V/\phi + 1$. There are a number of reasons why this is not realized in the actual case:

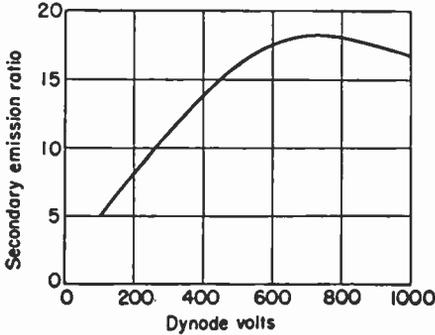


FIG. 4. When primary electron stream is directed normally to the surface of the dynode, the secondary emission ratio reaches a maximum value as the primary energy is increased. Curve is for magnesium oxide. (After Bruining and de Boer.)

1. The energies of the excited electrons are randomly distributed in direction; it is only the surface-directed component of the energy of an individual electron that determines whether it shall be emitted.

2. The majority of the emitted secondary electrons receive considerably more energy from the primaries than is required for emergence. Thus the average initial velocity of secondary electrons is greater than in the case of thermionic or photoelectric emission.

3. The energy of the primary electron, especially when the potential of the dynode with respect to the cathode is high, may not be expended until the primary electron has penetrated to a considerable distance below the surface

of the dynode, in which case the chance that electrons excited by the primary shall be emitted is considerably reduced.

Typical examples of the manner in which the depth of penetration of the primary electron affect the magnitude of the secondary electron current are shown in Figs. 4 and 5, both of which were redrawn from the data of Bruining and de Boer.

In Fig. 4 the ratio δ of secondary to primary current for magnesium oxide-coated platinum dynode is seen to rise from zero at very low dynode voltage almost linearly at low voltage, then gradually reaches a maximum value at about 700 volts, after which δ falls to lower values as the voltage is further increased. The curve is taken as a demonstration of the contention that, when the energy of the primary electron is high, much of it is transferred to other electrons at a distance so far below the surface that the chance of a secondary electron being emitted is greatly reduced.

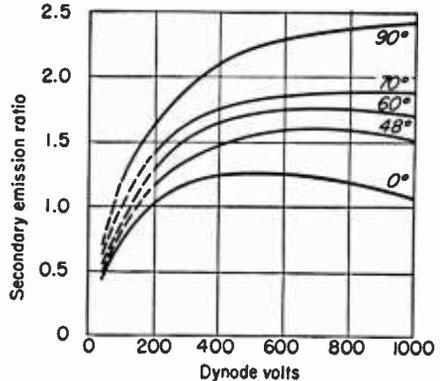


FIG. 5. Secondary emission ratio at various angles of incidence of primary electrons. Curve is for nickel. (After Bruining and de Boer.)

Figure 5 shows the manner in which δ varies with the angle of incidence of the primary beam when the magnitude of the primary current and the potential of the dynode are held constant. It is assumed that, as the angle between the beam and the surface of the dynode is reduced, the depth of penetration of the primary electrons is reduced, hence the average distance between the surface and the point at which the maximum excitation occurs is lowered. The result is an increase in the secondary electron current or δ .

In practice, dynodes are made by activating the surface of a base metal either by oxidation, by distillation onto it of a sensitizing material, or by heat-treatment to bring an impurity sensitizer to the surface.

An example of the first-named class of activation is that in which the base metal is an alloy consisting largely of magnesium, such as Dowmetal. The dynode is cleaned by mechanical processes, then heated in an atmosphere of oxygen from which water vapor and nitrogen have been excluded to produce a thin film of magnesium oxide.

In sensitizing by the distillation of sensitizing materials onto a base, one of several processes may be used. Barium or magnesium, or a combination of both, may be volatilized onto a nickel base *in vacuo*, then oxidized by admitting a small amount of oxygen to the device and applying a spark coil so that a glow discharge is created.

A method of more recent origin comprises using a well-cleaned and outgassed copper electrode upon which a film of cesium is deposited by vaporization from a side tube; the cesium is then oxidized by a glow discharge in low-pressure oxygen. After removal of the residual oxygen, a further small quantity of cesium is admitted. Dynodes made by this process have superior properties when properly prepared, the reason apparently being that the secondary electrons originate within the cesium oxide lattice where they are readily excited, while the metal cesium coating, because of its low work function, facilitates their escape.

For activation by means of an impurity material in the dynode, the dynode ordinarily consists of silver containing a few per cent by weight of magnesium. Activation consists in heating the dynode for a period of at least several minutes in the presence of oxygen, during which time some of the magnesium diffuses to the surface where it becomes oxidized.

In most cases where secondary emission is used, care is usually taken to ensure that the dynode is shielded from being contaminated by the cathode. While it is true that an electrode may be activated solely by contamination from an oxide-coated cathode, such activity is low compared to that produced by the methods just described and has relatively short life. A good dynode is, therefore, effectively spoiled by cathode contamination. The orbital-beam tube is an example of one way in which the dynode is protected.

Secondary emission often occurs as a spurious effect in various kinds of electron tubes, especially through transfer of activating material from the cathode to the anode or a grid. This is particularly noted in tubes employing oxide-coated cathodes. In the case of tetrodes the effect appears as a major shift in screen-grid current and in plate resistance during the life of the tube even with no change in cathode current. A fairly effective remedy for this condition is to spray the grid with carbon black or boron carbide and to carbonize the plate by first oxidizing the surface, then heating it in a carbon-bearing gas such as methane, CH_4 . The real solution to the problem is, of course, the employment of a suppressor grid held at a low voltage such that the field on both sides of it suppresses the emission of secondary electrons.

18. Field Emission. From an engineering standpoint, field emission is of less interest in the radio electron tube than are thermionic, photoelectric, and secondary emission currents, hence the discussion will be confined to a few remarks.

The relations describing quantitatively the phenomena of thermionic and photoelectric emission assume zero field at the surface of the cathode. When the field is not zero, the effect is equivalent to a change in the net work function by an amount

$$\Delta\phi = -\sqrt{-eF} \quad (17)$$

where F is the field or potential gradient just outside the surface of the cathode. In practical units,

$$\Delta\phi = -3.8 \times 10^{-4} \sqrt{-F} \quad \text{volts} \quad (18)$$

in which F is in volts per cm.

Field emission is of greater interest in gaseous discharge devices than in high-vacuum

tubes, but the phenomenon is one which must be considered in any application where strong fields are encountered, as in transmitting and rectifier tubes.

While in most high-vacuum tubes field emission is the cause of undesirable effects, it has been useful in some applications, *e.g.*, in certain television power-supply rectifiers. Here a diode contains an extremely fine tungsten filament coated to several times its bare weight with barium and strontium oxides. In operation, the filament is heated by the current passing through the thick coating, hence it functions as a thermionic cathode. In starting, however, the emission is first obtained by the strong field produced by the high voltage of the anode. Since the rectifier circuit is unloaded at the start, the voltage will be high and there will be a strong field at the filament. For a diode having axial symmetry, the field before conduction starts may be taken as $F = -V/r_0 \ln(r/r_0)$ where V is the anode potential, r and r_0 are, respectively, the anode and filament radii. As an example, if V is +20,000 volts with anode and filament radii taken as 0.5 and 0.001 cm, respectively, the field will be -3.22×10^6 volts per cm. By Eq. (18), the work function is lowered 0.68 volt. If the cathode has been properly processed to have a low work function, the field emission at room temperature is great enough to result in a rise in temperature, owing to the resistance of the coating, so that the current rises rapidly to the space-charge-limited value.

Spurious emission currents from grids and anode present certain problems which have been solved in several ways. A brief description of the most common practices will be given.

Grid emission may occur by any of the means already mentioned, the most common being thermionic or primary emission, and secondary emission; photoelectric emission is encountered in some tubes where the grid current due to gas or leakage is low enough for the photoelectric effect to become a factor.

Primary grid emission is encountered in receiving tubes employing oxide-coated cathodes where the spacing between grid and cathode is low and the dissipation on the electrodes is high in relation to their area. This is particularly noted in the output tubes, especially those intended for use in a-c-d-c receivers.

In most of the commercial receiving tubes where coated cathodes are used and where, without special precautions, grid emission would be a factor, radiators are welded to the ends of the grid side rods to suppress secondary electrons. Attention is given to the thermal conductivity both of the laterals and of the side rods; pure metals are preferred for the former, while beryllium-hardened copper is frequently used for the latter. For tubes used in oscillator service, the control grid is sometimes gold-plated or sprayed with finely powdered boron carbide if the type of operation is sufficiently severe to warrant this added expense.

For tubes employing thoriated tungsten filaments, the means employed to hold the grid emission to a low value depend to some extent upon the anticipated maximum grid temperatures. Where the temperature is not excessive, the grid may be made of platinum or of a more refractory wire which is platinum-clad, since platinum has the ability to absorb any thorium deposited upon it. For extremely high temperatures, the grid may be coated with certain refractory oxides, provided they do not break down with evolution of oxygen and cause decarburization of the filament. Good results have been reported with the oxides of zirconium and beryllium on molybdenum grids.

The means employed for the reduction of primary grid emission are usually effective in the control of secondary emission. However, the latter effect may be encountered in tubes having grids that operate at relatively low temperatures, in which case they may be sprayed with substances having a low coefficient of secondary emission such as carbon black, graphite, or zirconium.

Tubes are frequently designed for use wherein the anode may, during part of a cycle, be at a potential lower than some other electrode, as in rectifiers and tetrodes; hence either primary or secondary emission from the anode may be a serious factor unless precautions are taken to prevent or limit its occurrence. Where the tube is a diode rectifier, emphasis is placed upon the radiating properties of the anode so that the temperature will be low enough to limit primary emission. For small tubes,

carburized nickel or aluminum-clad steel anodes are preferred. In large tubes the anode is made with radiating fins and is sand-blasted and coated with zirconium, the latter serving both as a getter and in improving thermal radiation.

For tetrodes, the problem of thermionic emission from the anode is less frequently encountered. Secondary emission may be held at a minimum by the use of carburized metal as in the case of the old type 24A tube, or by a zirconium spray on the inside of the anode. To a considerable extent, however, this problem is more of academic than of practical importance, since the use of a suppressor grid as used in pentodes is a more practical solution.

SPACE-CHARGE-LIMITED CURRENTS

19. Diodes. A space-charge-limited current in an electron tube is an electron current which depends only upon the configuration and potentials of the electrodes and is *not* limited by the thermionic emission of the cathode. In this section all the relations are to be taken as being of practicable value only to the extent that adequate emission is available.

In a diode having a unipotential cathode, the current is given by the relation

$$i = Ge_p^{3/2} \tag{19}$$

in which G is the perveance and e_p the instantaneous anode voltage. In the formulas which follow, G will be expressed in terms of amperes per volt^{3/2}. The name of the unit has not been standardized.

Where the cathode is a filament, the voltage drop E_f along the filament must be considered, and this requires modification of Eq. (19). To be able to use the values of perveance given in succeeding paragraphs, the expressions are written in the following form for the two cases: $e_p \leq E_f, e_p \geq E_f$.

For $e_p \leq E_f$,

$$i = \frac{2}{5} E_f G e_p^{5/2} \tag{19a}$$

when the return connection is made to the negative end of the filament. It is to be noted that when $e_p < E_f$, an electron current is drawn from only that part of the filament which is negative with respect to the anode.

For $e_p \geq E_f$,

$$i = \frac{2}{5} G [e_p^{5/2} - (e_p - E_f)^{5/2}] \tag{19b}$$

when the return connection is made to the negative end of the filament.

Where the value of e_p is several times that of E_f , it is ordinarily of sufficient accuracy to use the relation

$$i = G(e_p - E_f/2)^{3/2} \tag{19c}$$

for connection to the negative end of the filament. For connection to the center tap of a filament-heating transformer, E_f in Eq. (19c) is omitted in considering the d-c component of the current.

For *parallel-plane* electrodes,¹

$$G = 2.34 \times 10^{-6} A/x^2 \tag{20}$$

where A = electrode area

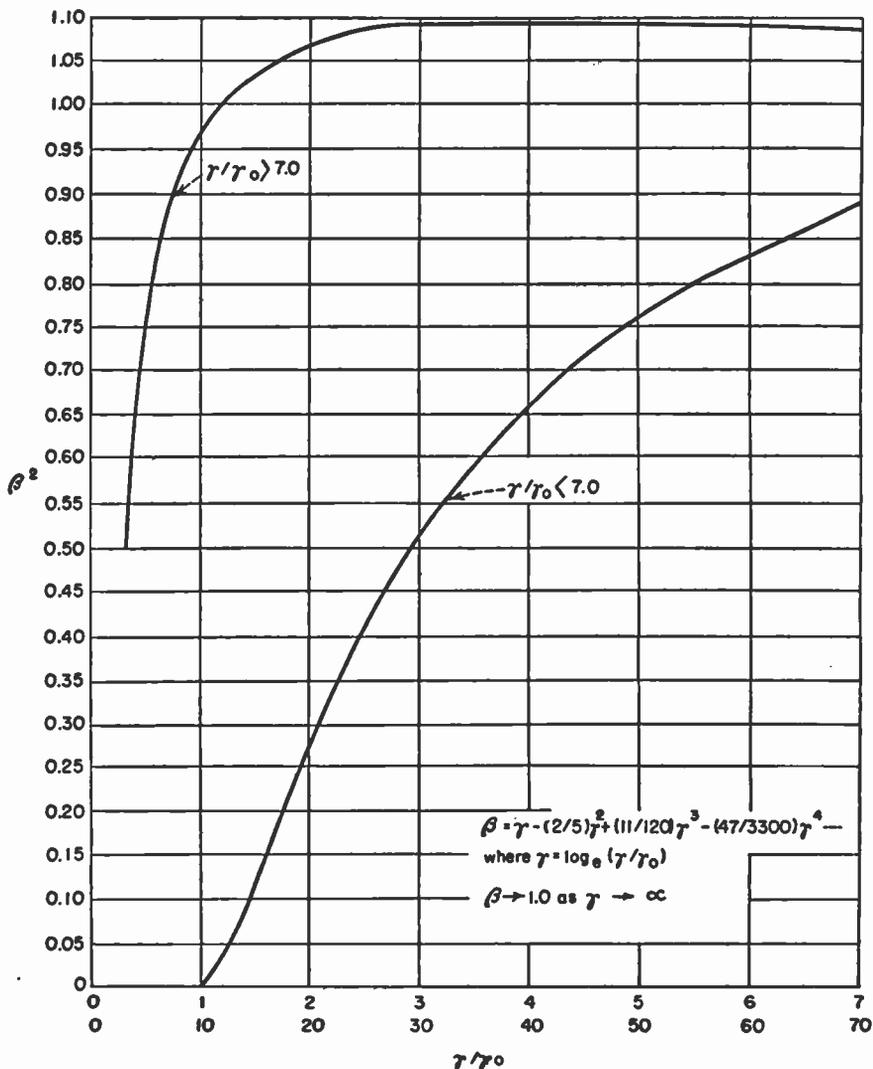
x = electrode spacing

The physical dimensions may be taken in any units for Eq. (20) or the formulas to follow, since they all turn out to be ratios.

¹ The perveance Eqs. (20) and (21) are to be taken as being accurate only under the hypothetical condition that the space-charge-limited current i is just equal to the temperature-limited emission current i_s . Where $i_s > i$, there is a retarding field at the surface of the cathode as shown in Fig. 9, p. 566, so that the magnitude of the perveance is increased as though the spacing s were reduced or r_0 were increased.

For parallel planes, the distance s is to be reduced by s_{min} , the distance between cathode and plane of potential minimum or virtual cathode. Where $i_s \gg i$, it is sufficiently accurate in most cases to take this distance as $s_{min} = 2.7 \times 10^{-6} T^{3/4} / i^{1/2}$ cm where T is the temperature of the cathode in deg K and i is the current density in amp per sq cm.

In the case of coaxial cylinders where r/r_0 is not great, the correction in r_0 is somewhat less than for planes and becomes vanishingly small for high values of r/r_0 .

FIG. 6. Values of β^2 for r/r_0 .

For coaxial cylinders,¹

$$G = 14.65 \times 10^{-6} L / r \beta^2 \quad (21)$$

where L = length of electrodes

r = radius of anode

β^2 = a function of r/r_0 shown graphically in Fig. 6, r_0 being the cathode radius

Figure 7 shows in cross section two types of diode structure that are of practical interest. Rigorous solutions for the perveance of such electrode configurations have

¹ See footnote on page 563.

not been published, but the formulas given on the figure are satisfactory for design purposes. These formulas are based on the transformation required to produce the same average charge density at the cathode as would be obtained at the same voltage by an equivalent cylindrical anode. The values of β_x^2 and β_y^2 are the ordinates of Fig. 6 when the abscissas are r_x/r_0 or r_y/r_0 .

For a cylindrical cathode and an anode of more-or-less oval cross section, the formula given in Fig. 7b is applicable, except that the equivalent cylindrical anode radius will be somewhat less than 1.31 times one-half the minor axis, depending upon the curvature. In practical design problems the proper factor is determined experimentally from preliminary tests based upon a reasonable guess. Although such a shape would rarely be used in a diode, the control grids of receiving tubes are frequently oval and

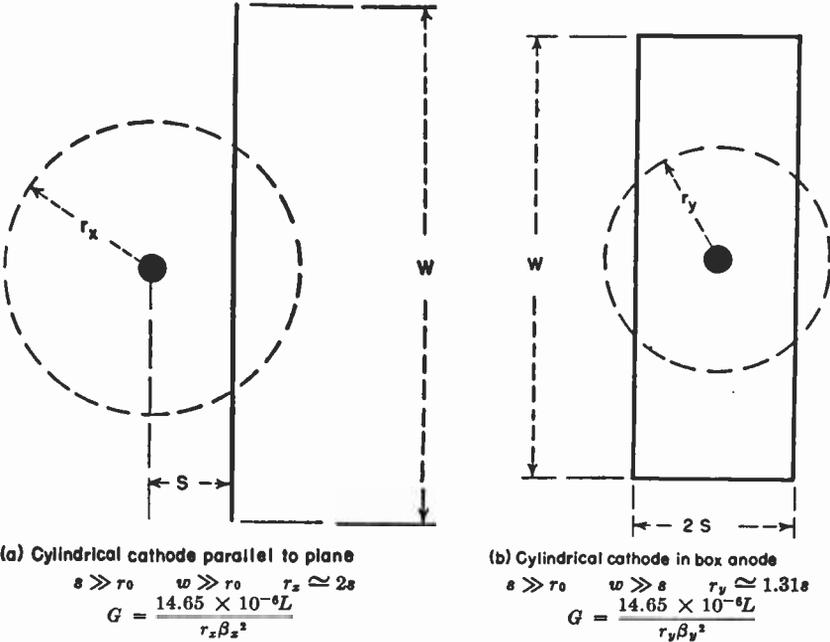


FIG. 7. Diodes which do not have axial symmetry. Cathode radius is r_0 , the active length is 1.

this method is applied in such cases. The manner in which the grid is replaced by an equivalent diode anode is noted in a subsequent paragraph.

The formulas for perveance all are based upon the assumption of zero initial electron velocity. Where voltages are high and the cathode-anode spacing large, these formulas are used in the simple forms given. However, at fairly low values of anode potential the spacing factors are interpreted with better accuracy when they are taken between anode and virtual cathode rather than between anode and the actual emitter. Although it is apt to be more a matter of academic than of practical interest to determine the location of a virtual cathode for a given set of conditions, it is well to bear in mind that the perveance formulas are apt to have little meaning when the anode-cathode spacing is less than a few thousandths of an inch and the voltage is very low.

20. Contact Potentials and Other Internal Effects. Where e_p of Eq. (19) is of the order of 10 volts or more, this can be taken as the instantaneous voltage e_0 applied at the terminals of the tube. At lower voltages, e_p must be taken to include certain inter-

nal effects, primarily the work functions of the electrode surfaces and initial electron velocity. As the voltage becomes very low, the current departs from the three-halves law of Eq. (19) and is said to be limited by a retarding field; usually some current is still obtained even when the applied voltage is slightly negative.

The contact potential difference between the electrodes is taken as minus the difference in the work functions ϕ_k and ϕ_a of their surfaces. Thus if there is no potential difference applied between the two electrodes, the potential just outside the anode is $-\phi_a$ volts with respect to the inside of either electrode, and the potential just outside the cathode is $-\phi_k$. So far as the field is concerned, it is the same as though the anode is placed at a potential $-(\phi_a - \phi_k)$ volts with respect to the cathode.

The principal effect of the initial electron velocity is the existence of the so-called *virtual cathode* at some distance from the cathode, the potential of which is less than any other place between the electrodes. For parallel planes, the potential of the virtual cathode with respect to the outside of the cathode is given as

$$V_{\min} = (-T/11,600) \ln (i_a/i) \quad (22)$$

where T = temperature of cathode

i = current to anode

i_a = temperature-limited emission

While the contact potential difference and the voltage V_{\min} are not the only internal correction voltages, together they constitute the important ones from an engineering standpoint, their sum being denoted by the symbol ϵ . Thus Eq. (19) may be taken as

$$i = Ge_p^{3/2} = G(e_b + \epsilon)^{3/2} \approx G[e_b - V_{\min} - (\phi_a - \phi_k)]^{3/2} \quad (23)$$

where e_b is the externally applied voltage and ϵ is the internal correction voltage, its value ordinarily being of the order of $1/2$ volt for small tubes when well aged.

Figure 8 shows graphically the manner in which ϕ_k , ϕ_a , V_{\min} , and e_b combine to control the space-charge-limited current in a tube. Taking the potential

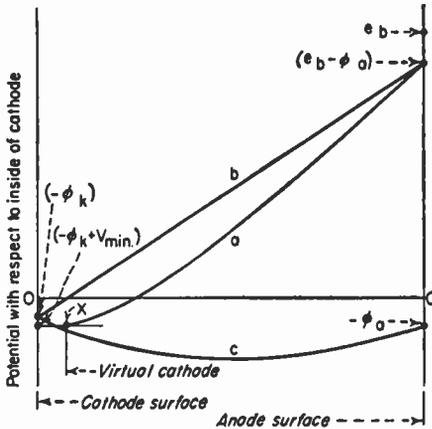


FIG. 8. Effects of work functions and potential minimum on the field in a diode with parallel plane electrodes. Curve *a* is the potential distribution for a space-charge-limited current; curve *b*, potential distribution when emission is very low; curve *c*, potential distribution with cathode hot but with zero applied voltage.

inside the cathode as zero, the potential just outside the cathode will be $-\phi_k$ volts, depending upon the material of which it is made. When the cathode is at a temperature high enough to supply a space-charge-limited current with an applied voltage e_b , the potential at all points is as shown by curve *a*. Owing to initial velocity, more electrons are emitted than reach the anode; their charge results in a depression in the curve at point *x* on curve *a*. The potential at point *x* is V_{\min} with respect to the outside of the cathode, or $(-\phi_k + V_{\min})$ with respect to the inside of the cathode. Now the *inside* of the anode is at a potential e_b with respect to the *inside* of the cathode; therefore, the *outside* of the anode is at a potential $(e_b - \phi_a)$ with respect to the inside of the cathode. The potential difference represented by e_p is here taken as that between the outside of the anode and the virtual cathode, hence $e_p = (e_b - \phi_a) - (0 - \phi_k - V_{\min})$ which is the term within the brackets of Eq. (23).

Line *b* of Fig. 8 shows the conditions existing when the cathode is cold; the potential minimum is zero, hence $V_{\min} = 0$ and $e_p = e_b + \phi_k - \phi_a$.

Curve *c* is for the case where the cathode is heated and $e_b = 0$; the potential difference between anode and cathode is $\phi_k - \phi_a$. Since i is now less than when $e_b > 0$,

V_{\min} (point x') is a greater negative voltage than before and is located much nearer to the anode.

The experimental value of ϵ may be determined to a useful degree of accuracy when the plate current and plate resistance are known. If it is assumed that the perveance G and potential minimum V_{\min} in Eq. (23) are relatively constant with change in applied voltage e_b , the derivative de_b/di may be taken as the plate resistance r_p of the tube at the corresponding plate voltage. The result is solved for ϵ , giving

$$\epsilon = \frac{3}{2}ir_p - e_b \quad \text{volts} \tag{24}$$

where i and r_p are expressed in amperes and ohms, respectively.

The importance of the correction voltage ϵ in connection with tubes containing grids will be shown in the discussion of the transconductance of triodes.

21. Triodes. In triodes the space-charge-limited cathode current is controlled in a manner analogous to that in the diode. In a triode, or a tube with more than one grid, the voltage of each electrode contributes to a composite controlling voltage of an equivalent diode anode, the dimensions of which are influenced by the dimensions of all of the actual electrodes except the cathode. The expression for cathode current is

$$i = G(e')^{3/2} \tag{25}$$

where G is the perveance of a diode having the same cathode and an anode of the same dimensions as the equivalent diode anode. The voltage e' is the composite controlling voltage.

For a triode having a high amplification factor, the equivalent diode anode may be taken as coinciding with the grid. Where such an assumption is not warranted, the following relations for locating the equivalent diode anode are generally of sufficient accuracy for design purposes.

1. *Plane electrodes,*

$$x' = (\mu x_1 + x_2)/(\mu + 1) \tag{26}$$

where x' = distance between cathode and equivalent diode-anode

x_1 = distance between cathode and grid plane measured to center of grid wire

x_2 = distance between grid plane and plate

μ = amplification factor¹

Where the cathode is a plane parallel to the grid, the perveance is that given by Eq. (20) when x' is substituted for x . Where the grid and plate are parallel planes and the cathode is a wire, x' is substituted in the formulas given in Fig. 7a or 7b.

As a singular case, the equivalent diode-anode of a triode in which all three electrodes are parallel planes is located almost precisely in the plane of the grid if the composite controlling electrode voltage is expressed in the form

$$e' = \frac{\mu e_c + e_b}{\mu + 1 + \frac{2}{3}(x_2/x_1)} + \epsilon \tag{27}$$

While Eq. (26) is apt to be preferred in the general case, Eq. (27) is of special use in the design of microwave tubes where the computation of electron transit time from cathode to grid is required.

2. *Cylindrical electrodes,* the radius r' of the equivalent diode-anode is given with reasonable accuracy by

$$r' = \log^{-1} [(\mu \log \rho_p + \log \rho_c)/(\mu + 1)] \tag{28}$$

where ρ_c and ρ_p are, respectively, the radii of the grid (to center of grid wire) and of the plate. The value of r' given by Eq. (28) is used directly in Eq. (21) for perveance.

3. Where the grids are *oval* and the plate is either oval or circular in cross section, certain approximation methods are useful. The distance between the equivalent

¹ For design purposes, the amplification factor used is the mu factor (μ_{p0k}), meaning the magnitude of the quotient of plate voltage change by grid voltage change for constant cathode current.

diode anode and cathode axis is taken in the vertical plane where the grid-cathode spacing is minimum, and computed as for plane electrodes. The perveance may then be computed to a reasonable degree of approximation as for the case of a diode of similar shape.

The composite controlling electrode voltage e' for triodes, except where Eq. (27) applies, is given by the relation

$$e' = \frac{\mu e_g + e_p}{\mu + 1} \quad (29)$$

where e_g and e_p are the grid and plate voltages, respectively. Where e' is of the order of 10 volts or more, it is generally satisfactory to take the electrode voltages applied to the tube terminals, *i.e.*, $e_g = e_c$, $e_p = e_b$ where e_c and e_b are, respectively, the applied grid and plate voltages.

Where e' is low, as in receiving tubes having a high amplification factor, e_g and e_p of Eq. (29) must be taken to mean that correction voltages are included. Although the contact potentials of grid and plate are usually not identical, it is convenient to disregard this fact, the error in so doing being negligibly small. In this case the composite controlling voltage may be written in terms of the applied voltage and a single correction voltage in the form

$$e' = \frac{\mu e_c + e_b}{\mu + 1} + \epsilon \quad (30)$$

22. Evaluation of Correction Factor ϵ . As will be shown later, the value of ϵ is a factor in the determination of the characteristics of the tube, hence means are required for its evaluation. For purposes of design, it is computed from static and dynamic readings. For purposes of quality control an approximate value is obtained by a simple measurement. Both methods are described below.

1. For design applications, the value of ϵ is readily found from a set of bridge readings made under the conditions for which ϵ is to be determined. In a triode, the cathode current is the same as in a diode when the equivalent diode-controlling voltage e' as given by Eq. (25) is substituted for e_p in Eq. (19). The expression becomes

$$i = G \left(\frac{\mu e_c + e_b}{\mu + 1} \right)^{3/2} \quad (31)$$

Now if the grid is negatively biased, the current i is the plate current i_b and the transconductance g_m is the rate of change in plate current with change in grid voltage, plate voltage held constant. Assuming the perveance G in Eq. (31) to be constant with change in grid voltage, the derivative di/de_c is taken and G eliminated. This results in the useful relation

$$\epsilon = \frac{3i\mu}{2g_m(\mu + 1)} - \frac{\mu e_c + e_b}{\mu + 1} \quad \text{volts} \quad (32)$$

Thus the measured values of plate current, amplification factor, and transconductance at the voltage conditions of measurement may be set into Eq. (32) to obtain ϵ . The value so obtained is in most cases slightly too high because G was taken to be constant, but the error usually cancels out when such values of ϵ are put back into a design equation where the same assumption is allowed. Attention is called to the fact that, if the current i is expressed in milliamperes, g_m is to be expressed in millimhos.

2. While Eq. (32) gives the value of ϵ with accuracy necessary for purposes of design or redesign, test results satisfactory for quality-control purposes may be obtained using the circuit shown in Fig. 9. The resistor R ordinarily has a value of 10 megohms, and the plate supply is adjusted to a value that results in a plate current of approximately rated value for the average tube when the grid return is through resistor R (switch in position *b*). The cathode is heated at rated operating voltage, and the plate current is carefully noted. The switch is then thrown to position *a* with the grid microammeter shorted and the plate current returned to its original value by adjustment of the potentiometer. The voltmeter then reads the internal

correction voltage ϵ , sometimes incorrectly referred to as the "contact potential." It should be noted that for directly heated cathodes the test should be made with direct current only and with the grid return made to the negative end of the filament.

The correction voltage is usually positive, but tubes are occasionally found in which the correction voltage is of negative polarity. Where knowledge of the magnitude of such a voltage is required, a microammeter may be placed in the grid circuit and the voltmeter reversed so that a positive voltage is applied to the grid. The voltage at which grid current just begins to flow is then noted. In either case, the correction voltage is minus the voltage being applied by the potentiometer.

In most cases where ϵ is of importance, *i.e.*, in high-gain receiving tubes, its value is likely to be between +0.3 and +1.0 volt. Where a value in excess of 1 volt is noted, the tube may be poorly aged. Further comments on the correction voltage will be found in the paragraphs dealing with transconductance and converter tubes.

23. Mu Factors. In general, the mu factor μ_{lmn} of an electron tube is the magnitude of the quotient of the alternating voltage of electrode *l* by the alternating voltage of electrode *m* under the condition that the current to electrode *n* remain constant. In amplifier circuits, the amplification factor μ_{ppp} is of most general interest, although, as has been noted, μ_{pjk} is used in tube design, since it is the total cathode current that is directly related to the composite controlling voltage.

In certain cases of restricted interest the mu factor μ_{ppg} of a triode with positive grid and negative plate is taken, the value of the factor being less than unity. It should be pointed out that in such operation the perveance *G* is considerably reduced by electrons that pass twice through the grid, the reason being that an electron moving in the grid-to-cathode direction has the same effect upon the space charge as an electron moving in the usual direction.

In the paragraphs immediately following, it is assumed that $\mu = \mu_{pjk}$ or, of course, μ can be μ_{ppp} if it is assumed that the grid draws no current from the electron stream and that the frequency is low.

Several formulas have been published for μ in terms of the dimensions of the tube. Those of Vogdes and Elder are commonly used and are given below. The distance between grid and cathode and the shape of the cathode are not factors, although they do have some influence when the spacing or the composite controlling electrode voltage is low.

For parallel-plane electrodes,

$$\mu = k_1 x_2 - k_2 \tag{33}$$

For coaxial cylinders,

$$\mu = k_1 \rho_g \ln (\rho_p / \rho_g) - k_2 \tag{34}$$

where x_2 = grid-plate spacing for parallel electrodes

ρ_g = radius of grid

ρ_p = radius of plate

$k_1 = 2\pi n / \ln \coth 2\pi nr$

n = grid turns per unit length

r = radius of grid wire

$k_2 = (\ln \cosh 2\pi nr) / (\ln \coth 2\pi nr)$

It is seen that, for plane electrodes, μ varies directly as the grid-plate spacing. For cylindrical electrodes, differentiation of μ with respect to ρ_g shows that μ is maxi-

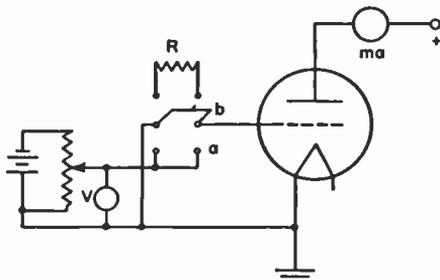
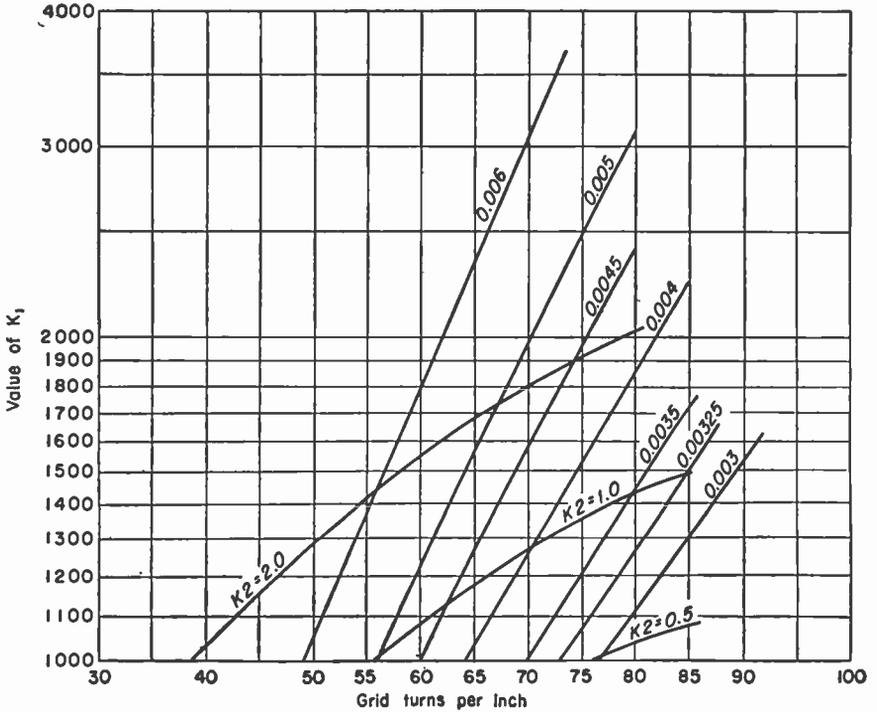
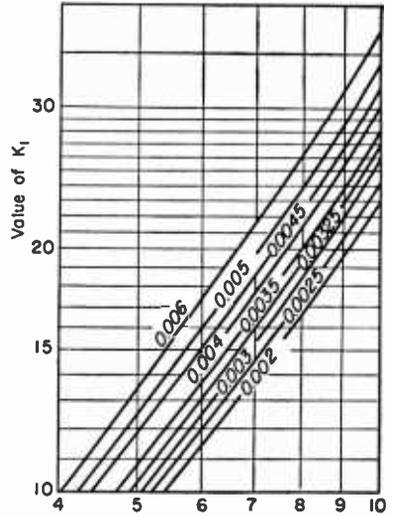
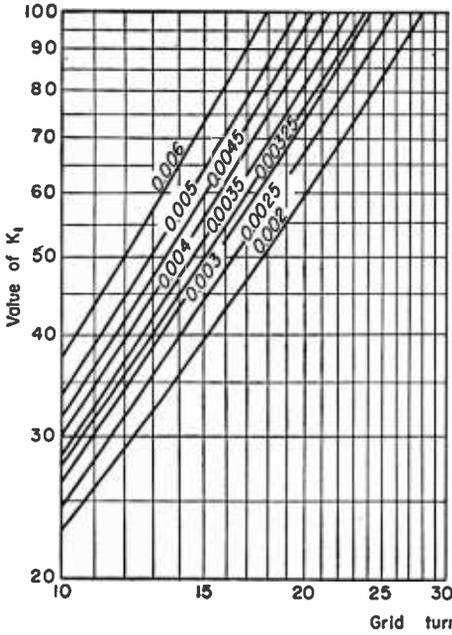


FIG. 9. Circuit for rapid check of grid internal correction voltage. A microammeter may be placed in the grid circuit for checking both positive and negative values by the slower method.



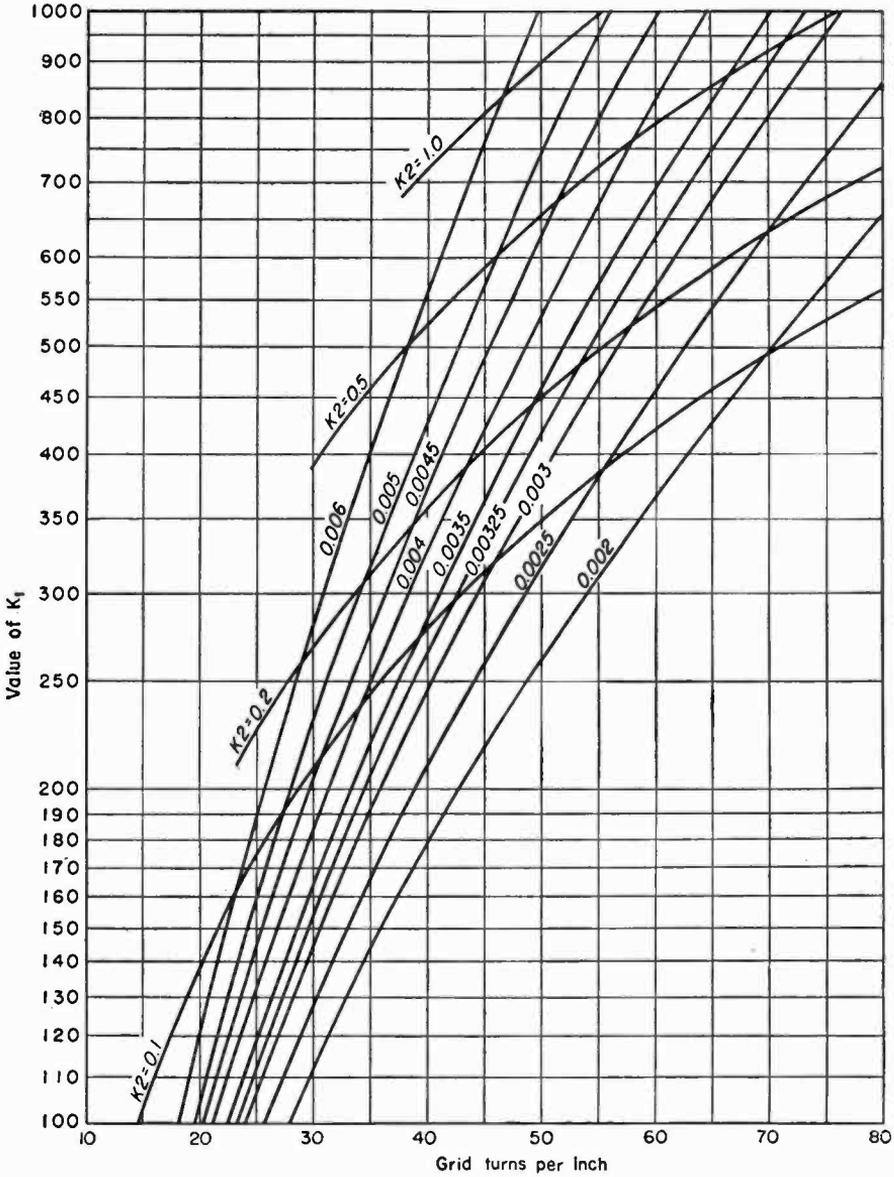


FIG. 10 (pages 570 and 571). Evaluation of k_1 in terms of wire size in inches and winding in grid turns per inch.

imum when $\rho_g = \rho_p/2.718$, which means that, for this ratio, accidental variations in ρ_g or ρ_p result in minimum variation in μ .

Values of k_2 are seen not to vary with x_2 , ρ_g , or ρ_p and are small in magnitude, hence they are frequently ignored in adjusting the dimensions of a tube to achieve an alteration in μ . Figure 10 provides a convenient means for the evaluation of k_1 in terms of wire size in inches and winding in turns per inch.

It is often necessary to design triodes in such a manner that the distance between the grid and anode is varied circumferentially, *e.g.*, where the grid is oval and the anode either cylindrical or rectangular in cross section. It is evident that the density of the electron stream is likely to be greatest where the grid-plate spacing is smallest, and so that part of the structure will largely determine the amplification factor. For such a tube, a preliminary design is tested in which the minimum distance is taken as x' in Eq. (26); final adjustment will be based upon an effective value of x' determined by measurement of μ for the preliminary samples.

24. Mu-slop and Cutoff. When the anode current is reduced either by decrease in anode voltage or increase in negative grid voltage, μ is observed to decrease. This decrease is generally not great where the effective controlling electrode voltage is of the order of several volts, but as the current is decreased to the order of some microamperes the decrease in μ becomes pronounced. This effect, known as mu-slop, is due to a number of causes. Slight variations in dimensions, such as the misplacement of a grid wire or incomplete coverage of the plate by the grid, may permit the passage of current through one part of the tube when it has been cut off at all other parts. The effect is that the μ of the tube is, under this condition, the same as though the entire tube were made with a more open-spaced grid.

Mu-slop plays an important part in the determination of cutoff, although the extent is not always easy to determine. Reference to Eq. (31) indicates that the current is zero when the bracketed term is zero. However, as the term approaches zero, μ decreases and ϵ increases,¹ neither change being readily predictable. Mu-slop is most pronounced in receiving tubes having indirectly heated cathodes while the change in ϵ is greatest for tubes having high emission. Furthermore, the mu-slop at a given value of plate current increases as the plate voltage is increased. In view of these uncertainties, it is not possible to predict the precise value of grid voltage at which the current shall be some arbitrary low value. However, the following empirical relation is used in setting up a quality test for cutoff voltage for a current maximum of 10 μa :

$$e_{c0} \cong -1.35e_b/\mu \quad (35)$$

where e_{c0} is the grid voltage applied, e_b and μ are the rated values. For pentodes, e_b is replaced by e_{c2} and μ by the triode μ_1 .

25. Transconductance and Plate Resistance. Differentiation of Eq. (31) also provides the relation which gives the grid-plate transconductance or mutual conductance g_m in terms of amplification factor, plate current, and electrode voltages. Equation (31) is to be taken with e_c sufficiently negative so that all of the current goes to the plate. The result may be expressed in the form

$$g_m = \left(\frac{di_p}{de_c}\right)_{e_b} = \frac{3}{2}G \left(\frac{\mu}{\mu+1}\right) \left(\frac{\mu e_c + e_b}{\mu+1} + \epsilon\right)^{1/2} \quad (36)$$

where i_p is the plate current, the grid current being zero. Equation (36), with elimination of G by means of Eq. (31), gives the useful relation

$$g_m = \frac{3}{2} \frac{\mu i_p}{\mu e_c + e_b + \epsilon(\mu+1)} \quad (37)$$

from which Eq. (32) was taken by rearrangement of the terms. Note that if i_p is expressed in milliamperes, g_m is in millimhos.

¹ The potential minimum, an important part of ϵ , is seen in Eq. (22) to vary as i varies; as i decreases, V_{min} increases negatively while the contact potential difference $-(\phi_a - \phi_k)$ presumably remains constant, thus ϵ changes in the positive direction.

In the discussion of diodes it was indicated that the perveance formulas do not hold when the anode-cathode distance becomes very small. This is also true in tubes having grids, so that Eqs. (36) and (37) also fall down under such conditions. The maximum theoretical value of g_m is of the order of $10^4 \mu\text{mhos per ma.}$

Expressions for the plate resistance r_p may be derived just as Eqs. (36) and (37) were obtained by taking $1/r_p = (di_p/de_b)_{ec}$, the result being merely a demonstration of the well-known rule that

$$r_p = \mu/g_m \quad (38)$$

A relation that is very useful in tube design is obtained from Eq. (37) for determining the amplification factor necessary when the values of g_m , i_b , e_c , and e_b are specified, which constitutes one of the first steps necessary in the design of any tube having one or more grids. The equation is

$$\mu = \frac{2g_m(e_b + \epsilon)}{3i_b - 2g_m(e_c + \epsilon)} \quad (39)$$

Inspection of Eq. (39) shows that, for required values of g_m , i_b , e_c , and e_b , a high value of ϵ requires a higher value of μ than would be required when ϵ is low. It is, therefore, important to know at the start, especially in a high- μ triode or a high-gain, sharp cutoff pentode, how low a value of ϵ may be expected from the tube when well aged or what its value will be after some hours of operation.

Control of ϵ . As previously noted, the expected value of ϵ cannot be predicted with accuracy, but reasonable estimates of the probable value can be based on similar tubes. Variations from average for a type, whether the variation is among individual tubes or changes during life, are important considerations in the control of quality in high-gain tubes.

Equation (32) assumes full space-charge-limited current, but when the current of a tube is partially temperature-limited, the quotient of plate current by transconductance is high, hence such a tube may be expected to show a very high value of ϵ when so computed. For this reason, tubes which apparently have an internal correction voltage much above 1 volt should be checked for emission.

Apart from the effect of poor or unstable emission, the principal cause of variation in ϵ is usually the variation in contact potential of the control grid. In tubes having coated cathodes this variation may be as much as 1 volt; in tungsten-filament tubes it is usually small. Where a coated cathode is used, the grids appear to become contaminated by the cathode during life so that the actual contact potential difference approaches a very low value, hence ϵ approaches the value of $-V_{\text{min}}$. In most indirectly heated tubes, ϵ will become stable after the tube has been in use for 50 or 100 hr. In high- μ triodes and in high-gain receiving-type pentodes with coated cathodes the grid contact potential may be partially stabilized by taking tubes initially at room temperature and raising the heater voltage fairly rapidly to about twice the rated value and holding the condition for about 30 sec, then turning off the current. No other voltages are applied and the cathode must not be heated longer than indicated, since otherwise the grid becomes heated and the condition is reversed. The process may be repeated a number of times, usually with some continued change in ϵ , provided sufficient time is allowed between treatments for the grid to cool off. The treatment is the same regardless of whether ϵ is originally higher or lower than the stable value finally obtained.

In tubes having coated cathodes made for industrial purposes, where constant characteristics over a period of some thousands of hours is a primary requisite, special means are employed for maintaining the contact potential at its initial value. The cathodes are frequently designed so that the temperature is somewhat lower than that of similar radio receiving tubes. This reduces the rate at which coating is volatilized, hence the maintenance of the coated diameter and activity of the cathode is improved and the rate of grid contamination is reduced. The control grids are usually silver plated so that any barium deposited may be absorbed by the silver.

26. Division of Current in Positive-grid Triode. The manner in which the total cathode current of a positively biased tube divides between grid and anode cannot be specified with any high degree of accuracy. There are, however, formulas that yield fair results in the absence of grid primary or secondary emission and are based on the following concepts:

1. When the grid is at its "natural" potential, the ratio of grid current to plate current is the quotient of grid wire diameter by the distance between grid wires.

2. The ratio of plate current to grid current varies approximately as the square root of the ratio of plate voltage to grid voltage.

The term *natural potential* means the voltage E_γ which, applied to the grid when the plate is at the arbitrary voltage E_π , results in the same cathode current as would have been obtained with the grid physically removed. If the *cathode, grid, and plate are all planes*,

$$E_\gamma = E_\pi \left(\frac{x_1}{x_1 + x_2} \right)^{3/2} \quad (40)$$

where x_1 and x_2 are, respectively, the cathode-grid and grid-plate spacing.

When cathode, grid, and plate are all coaxial cylinders, the natural potential of the grid is

$$E_\gamma = \left(\frac{\rho_g \beta_g^2}{\rho_p \beta_p^2} \right)^{3/2} E_\pi \quad (41)$$

where ρ_g and ρ_p are, respectively, the grid and plate radii, β_g^2 and β_p^2 are obtained from Fig. 7 (see p. 565) using ρ_g/r_0 and ρ_p/r_0 , respectively, r_0 being the cathode radius.

For these two cases, one may then take as an approximation, in the absence of primary or secondary emission from the grid, the empirical relation:

$$\frac{i_b}{i_c} \approx \left(\frac{s-d}{d} \right) \left(\frac{e_b E_\gamma}{e_c E_\pi} \right)^{1/2} \quad (42)$$

where i_b and i_c = plate and grid currents

e_b and e_c = plate and grid voltages

s = spacing between grid wires on centers

d = grid wire diameter

For the case in which the cathode is a wire or ribbon, grid and plate are planes, Spangenberg gives the relation:

$$\frac{i_b}{i_c} = \frac{L + M(E_b/E_c)}{P - Q(E_b/E_c)} \quad (43)$$

where $L = \pi s p \mu - s \mu (p + f) D - 2 \pi p r_0 \mu$

$M = \pi s p - 2 \pi p r_0 + s f D \mu$

$P = s(p + f) \mu + 2 \pi p r_0 \mu$

$Q = s f D \mu - 2 \pi p r_0$

s = spacing between grid wires

r_0 = grid wire radius

p = grid-plate spacing

f = filament-grid spacing

μ = amplification factor of tube

$D = r_0 / 2f \ln(4ef/r_0)$, where $e = 2.718$

27. Negative-anode Triodes. Triodes are occasionally used in circuits wherein the anode voltage is made negative, the grid positive; the Barkhausen-Kurtz oscillator is such a device. At low frequencies, tubes may be employed with the output circuit between cathode and positive grid, the input between cathode and negative plate. In such a case the output resistance and gain may be made quite low; but where such a circuit is used, the gain is ordinarily of minor interest.

The *perveance* of the negative-anode triode is *no longer* that given as the quotient of cathode current by the three-halves power of the effective diode voltage; in fact,

perveance has very little meaning in such a case. The space-charge equations are based upon Poisson's equation which may be interpreted to say that the electrons in the space between cathode and anode are distributed in a certain fashion, *regardless* of the direction of their motion. When electrons are repelled by the plate and again pass between the grid wire, they tend to suppress the flow of newly emitted electrons. This is seen graphically in Fig. 11, where the cathode current is depressed by a negative bias on the suppressor grid. Suppose the rate at which electrons leave the cathode is the current I_0 . A fraction of these electrons is captured by the grid and the balance, $(1 - a)I_0$ return to the cathode. Then

$$I_0 + (1 - a)I_0 = I = GV^{3/2} \tag{44}$$

where I is the current when $a = 1$; *i.e.*, when no electrons are returned to the cathode and G is the perveance as ordinarily understood. The grid current of the negative-plate triode is, then,

$$i_g = aI_0 = \left(\frac{a}{2 - a}\right) I \tag{45}$$

28. Tetrodes. Tubes having two grids between cathode and anode are the more common types of tetrodes. The grid nearest the cathode, designated g_1 , is ordinarily operated with negative bias and the second grid, g_2 , is used to provide increased gain or power output and reduced control grid-plate capacitance. Such tubes suffer by comparison with pentodes in their performance, owing largely to the fact that secondary emission reduces the plate resistance r_p , therefore the gain; the secondary emission varies with life, causing a shift in tube characteristics.

The perveance G and internal correction voltage ϵ of a tetrode may be computed or measured provided the conditions are such that electrons are not made to reverse their directions and affect the space charge in the region of the cathode. Under ordinary conditions a small change in plate voltage produces a relatively insignificant change in cathode current. The change in plate current with small change in plate voltage is usually found to be accompanied by a change in screen-grid current which is opposite in direction and almost equal in magnitude, while the change in cathode current is very small indeed. Hence the perveance is practically the same as if the screen grid were replaced by an anode, the potential of this anode being that of the tetrode screen grid. Under these conditions,

$$i = i_b + i_{c2} = G \left(\frac{\mu_1 e_{c1} + e_{c2}}{\mu_1 + 1} + \epsilon \right)^{3/2} \tag{46}$$

where μ_1 = triode amplification factor

i_b = plate current of tetrode

i_{c2} = screen-grid current

e_{c1} = control-grid voltage

e_{c2} = screen-grid voltage

The value of G is then related to dimensions as in the case of a triode. For determining the perveance and the internal correction voltage of an experimental tube from bridge measurements, the screen grid and plate are connected at the socket and the normal screen-grid voltage is applied. The readings are then substituted in Eq. (46), $(i_b + i_{c2})$ now being the total triode plate current. Similarly, ϵ is obtained when the same data are placed in Eq. (32), the value of g_m now being that obtained for the triode connection and e_b being the voltage applied to the screen grid and plate tied together.

The transconductance of the tetrode is related to that of the triode as in Eq. (47) below, assuming that there will be zero change in the proportional division of current between screen grid and plate when a signal is applied to the control grid, and that the cathode current is the same for either triode or tetrode connection.

$$g_{m(\text{tetrode})} = (i_b/i_k) \times g_{m(\text{triode})} \tag{47}$$

where i_b = tetrode plate current

i_k = total cathode current ($i_b + i_{c2}$)

The relation $g_m = \mu/r_p$ holds for a tetrode as for any other tube. However, the values of μ and r_p are not readily predictable from the dimensions in view of the fact that the ratio of screen current to plate current changes when a signal is applied to the plate. This change acts to reduce r_p and μ . If the change were zero, the tetrode amplification factor would be approximately equal to the sum and product of (1) the μ of the tube as a triode with the screen grid omitted and (2) the μ of the tube as a triode if the screen grid were the only grid in it. To keep the amplification factor as high as possible, the distance between screen grid and plate is made relatively large and the screen grid may be treated to hold the secondary emission low, as by spraying it with graphite. In transmitting tubes the wires of the screen grid are sometimes placed directly behind the wires of the control grid, thus achieving an increase in plate resistance and reduction in screen-grid dissipation.

29. Pentodes. A pentode is an electron tube having five electrodes: usually a cathode, an anode, and three grids. This discussion will be limited to the case where the electrons from the cathode pass successively through all three grids to the anode. In practically all cases the pentode may be considered as a tetrode in which a third grid is added between screen grid and plate for the purpose of suppressing secondary emission; hence the third grid is often called a "suppressor" grid. The principal results of suppressing secondary electrons are an increase in the amplification factor (hence an increase in gain), increase in power output, and decrease in distortion for large-signal tubes.

The manner in which the third grid suppresses the flow of secondary electrons is readily understood. With the suppressor at zero potential while both screen grid and plate are positive, the field toward the suppressor is retarding for electrons on both sides. Since the average initial velocity of the secondary electrons is equivalent to only a few electron volts at most, the energy is generally too low to allow them to pass through the suppressor. To a certain extent, the control of the secondary electrons is analogous to cutoff in a triode. For example, if μ_3 is the suppressor grid-plate μ , one could expect the flow of secondary electrons from the screen grid to the plate to be blocked when $(e_b - e_{c2}) + \mu_3(-e_{c2}) = 0$. In the practical case μ_3 is made considerably greater, a minimum figure being

$$\mu_3 = 2(e_b - e_{c2})/e_{c2} \quad (48)$$

in which e_b is to be taken as the highest value of plate voltage likely to be encountered under full-signal operation. For ordinary receiving-type pentodes, μ_3 is about 15 to 20 when measured at $e_{c2} = 100$, $e_b = 250$, $e_c = 0$.

Figure 11 shows a typical curve of the electrode currents in a pentode plotted against suppressor voltage. Although the effect of the suppressor-grid voltage on the anode current is usually of no importance in ordinary amplifier operation, it is of interest in certain circuits such as multivibrators and square-wave generators. It is noted that the i_b-e_{c3} curve has a flexion point at which the suppressor grid-plate transconductance is maximum. This corresponds approximately to the conditions under which half of the electrons moving toward the plate are turned back by the suppressor grid. Under conditions approaching perfect symmetry, including alignment of the control and screen grids and a high value of μ_3 , the i_b-e_{c3} curve becomes very steep at the inflection point—which will occur with a very small negative bias on the suppressor grid—and the suppressor grid-plate transconductance approaches a theoretical maximum value which is of the order of 10,000 μ mhos per ma plate current.

In Fig. 11 the cathode current has been plotted against suppressor-grid voltage. It is seen that, where the plate current increases rapidly with decrease in suppressor bias, the total cathode current also rises. This may be interpreted as showing the manner in which the cathode current is limited by electrons returned toward it from beyond the control grid, or, this being granted, as showing that some of the electrons repelled by a bias on the suppressor grid may again pass through both screen and control grids. Figure 11 was taken from a tube having substantially cylindrical

electrodes. Had the filament been a wire and the other electrodes flat, the effect of the suppressor-grid bias on the cathode current would have been less noticeable.

The perveance of a pentode is taken in the same manner as in the case of a tetrode, the suppressor—where it has an external connection—being connected to the cathode. The equivalent diode-anode is then just beyond the control grid as in the triode, so that the data given in the discussion of diodes and triodes will apply here.

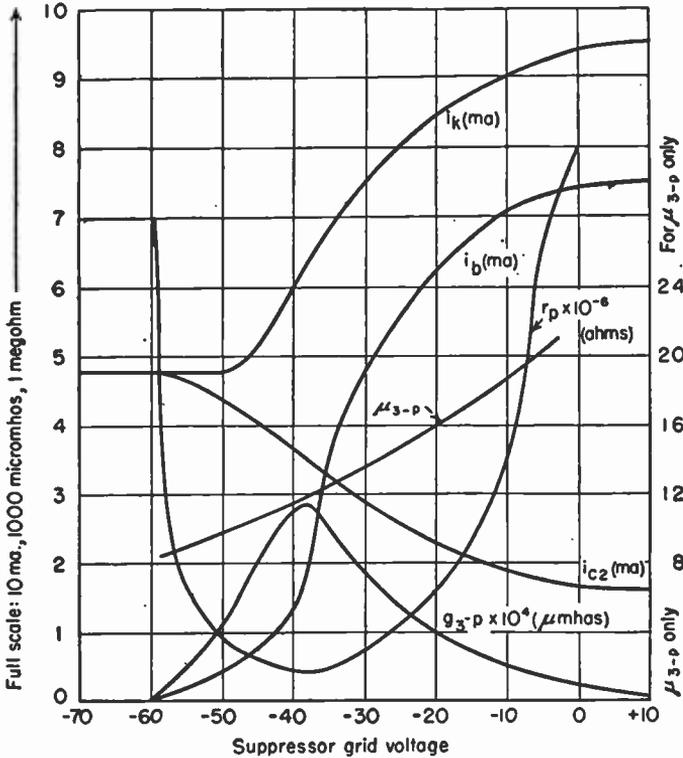


FIG. 11. Variation in i_b , i_{c2} , i_k , μ_{3-p} , and r_p with variation of suppressor-grid voltage.

Similarly, the equivalent controlling electrode voltage may be taken by considering the screen grid to be equivalent to a triode plate; hence,

$$e' \approx \frac{\mu_1 e_{c1} + e_{c2}}{\mu_1 + 1} + \epsilon \tag{49}$$

30. Pentode Design. While it is beyond the scope of a handbook to outline in detail the design of a pentode, mention will be made of a few points which may aid in an understanding of the more difficult steps. It is assumed that the objective is a tube having specified values of i_b , i_{c2} , and g_m at specified values of e_{c1} , e_{c2} , and e_b .

1. The first step is to decide on a value of the internal correction voltage. If the tube is to resemble an existing tube to the extent that the values of ϵ could be expected to be alike, the existing tubes can be read as triodes and ϵ computed from Eq. (32). In the absence of such data, ϵ can be taken arbitrarily as, say, 0.7 volt in the preliminary test.

2. The required triode μ is computed from Eq. (39), using e_{c2} in place of e_b and

e_{c1} in place of e_c . (Note that the screen current does not enter the relation, since the pentode g_m and i_b are the same fractions of the triode g_m and i_b , respectively.)

3. Assuming that the general configuration of the electrodes and the dimensions of the cathode have been picked, the dimensions of the composite anode are chosen to give the proper perveance as described in the paragraphs on diodes. The value of the perveance required will be

$$G = \frac{i_b + i_{c2}}{\left(\frac{\mu_1 e_{c1} + e_{c2}}{\mu_1 + 1} + e \right)^{3/2}} \quad (50)$$

where μ_1 is the required triode amplification factor as computed by the method outlined in the preceding paragraph.

4. The surface dimensions of the control grid are first taken as being those of the composite anode, and the surface dimensions of the screen grid then chosen to have reasonable clearance and shape. The exact surface dimensions of the control grid may then be determined as for a triode. See Eqs. (26), (28), (33), (34), and accompanying text. When these results are at hand, good judgment will frequently indicate that a change may be necessary in some of the parts arbitrarily chosen; e.g., a larger cathode might better be used, or a screen grid of smaller diameter might have some advantage. In the former case, a new composite anode must be computed; in the latter case, a new control grid will be computed for the smaller screen grid.

5. No attempt is made in the first design to produce a specified screen-grid current. The first batch of tubes must, in any case, be regarded as means to provide a basis on which to make corrections, and such corrections may include the correction for screen-grid current. In a shadow-grid (grid turns aligned) type of tube, the screen-grid current changes in the same direction as the diameter of the wire with which it is wound and, other factors being the same, increases as the distance between g_1 and g_2 is increased.

6. The dimensions of the plate are not especially critical so far as the design is concerned, but this is not to be taken as meaning that plates for a given design can be allowed to vary over wide limits. The plate is to be shorter than the grids to prevent the field at the ends from extending around the grids and influencing the electron current or resulting in unnecessarily high control grid-plate capacitance.

7. In view of Eq. (48) the μ of suppressor to plate must be adequate. On the other hand, μ_s must not be too high for the reason that the plate resistance of the tube may be too low, especially when operating under the condition that $e_{c2} \simeq e_b$. The reason for this effect is discovered when it is considered that electrons, in passing close to the screen-grid wires, may have their paths bent in such a way that, while their total momentum is unchanged, the component in the direction of the retarding field of the suppressor is reduced to the point where a slight difference in plate voltage will have a major effect upon the probability that such an electron shall reach the plate. The writer knows of no formula that accurately predicts the dimensions of the ideal suppressor grid. When the suppressor grid has an external connection, some check may be obtained by measuring the plate resistance r_p of the tube with the potential of the suppressor changed a few volts above and below zero; if maximum r_p is obtained with the suppressor negative, it is generally an indication that it would be advantageous to wind it with a tighter pitch; if r_p is maximum with the suppressor positive, the pitch may be opened. In any event, the final pitch selected is the result of comparative tests.

31. Shielding. In a pentode intended for r-f operation, shielding is used primarily for the reduction of control grid-plate capacitance. The design of shielding elements in a tube cannot be said to be based on formula. The evolution of shielding since the appearance of the first shielded tetrodes about 1925 might be considered as an example of intelligent cut-and-try engineering, particularly as exemplified in some of the high-perveance, single-ended pentode receiving tubes. In constructing experimental tubes

the engineer who does not have available the advice of persons having experience in this phase of tube design will probably obtain best results if control grid and plate leads are taken out at opposite ends of the tube and the design features of commercial tubes, so far as shielding is concerned, are carefully copied. In such cases, the shields of discarded tubes are often found to be worth salvaging for experimentation.

32. Remote-cutoff Tubes. Ordinarily, remote-cutoff tubes are designed to fit a preferred curve of control-grid voltage plotted against either transconductance g_m or conversion transconductance g_c .

The control grid may be wound with a continuously variable pitch, the most open part being in the center. It is more practicable, however, in pentode amplifiers to use only three pitches—the most open being in the middle, then one or a few turns of an intermediate pitch each side of center, and a constant, close pitch over the balance of the length of the grid.

So far as the pentode is concerned, an approach to the problem of design is made by considering that the tube has three parts: one part having a high value of triode μ , one part having an intermediate value, and a third part having a low value. The length of each part of the tube is the same as that of the corresponding parts of the control grid. In the five paragraphs immediately following, the problem is considerably oversimplified, although it does show in outline a means for setting up a preliminary test. Final design is achieved, as with other tubes, by adjustment of factors that may be computed by measurement of the preliminary samples.

For the pentode amplifier, suppose three points on the g_m - e_{c1} curve are fixed at the bogey values that the design is expected to achieve. Let $g_{m(1)}$ be the mutual conductance at the low-bias end of the curve, the control-grid voltage being $e_{c(1)}$; at an intermediate point the values are $g_{m(2)}$ and $e_{c(2)}$ while at a point near cutoff they are $g_{m(3)}$ and $e_{c(3)}$. Correspondingly, there will be three values of plate current: $i_{(1)}$, $i_{(2)}$, and $i_{(3)}$.

At grid bias e_{c3} , the tube will be cut off all except at the one turn in the center. Here the triode μ , $g_{m(3)}$ and $i_{(3)}$, and the voltages will be related as in an ordinary triode or pentode, and the pitch of the control grid, n turns per unit length, is computed. The length of this part of the tube is $1/n$, and its current and transconductance for this section may then be computed to a reasonable degree of accuracy at any other value of grid bias.

At the intermediate point the current and transconductance of the low- μ section are computed and subtracted from the required values. The difference between this and $i_{(2)}$ and $g_{m(2)}$ is then made up by computing the length and μ of the intermediate section to give a total current and transconductance of the proper value for this point.

Similarly, at the low-bias end, the current and transconductance of the low and intermediate sections are computed and subtracted from the required values $i_{(1)}$ and $g_{m(1)}$. The difference is then taken and the high- μ section computed to bring the totals to the required values.

Such a procedure will often result, on first trial, in a computed cathode length which is not satisfactory as a practical design. The perveance per unit length may then be raised or lowered, as by altering the surface dimensions of the control grid, and a new set of computations made.

33. Beam-power Tubes. Also classed as pentodes are the so-called beam-power output tubes employing a cathode, anode, two grids, and a pair of beam-confining plates, the latter taking the place of the suppressor grid in the high-gain pentode. The word *beam* is used to indicate that the wires of g_1 are interposed between the cathode and the wires of g_2 so that the electron stream is in the form of beams passing toward the anode. Figure 12 shows in cross section the electrode arrangement of a typical tube of this variety.

In a beam-power output tube the beam-confining plates are ordinarily connected internally to the cathode. While the relations between voltage, current, and mutual conductance and the dimensions of cathode, g_1 and g_2 are as in any other tube, the design from g_2 outward is still largely a matter of cut and try. Achievement of a satisfactory design is not difficult when the problem is understood in a qualitative way.

Under full-signal conditions the instantaneous plate voltage of the tube will reach a value much higher than that of the screen grid, and to prevent secondary electrons from g_2 reaching the plate, a retarding field must be provided. Reference to Fig. 12 readily discloses that the retarding field due to the beam-confining plates would be minimum midway between the edges, increasing toward the edges. Therefore, to have sufficient field midway between the edges of the beam plates and not too strong a field toward the edge of the electron beam, the clearance between g_2 and anode should be relatively large and the beam-confining plates placed about halfway between.

When the plate voltage swings to its minimum value, the field between screen grid and plate must also be retarding to prevent the passage of electrons from the plate back to the grid. This adds up to an indicated requirement that the beam-confining plates have a fairly generous clearance on both sides. It will be understood that when a tube is delivering power to a load, the plate current is maximum when the plate voltage is minimum, and that the electrons passing between the edges of the beam-

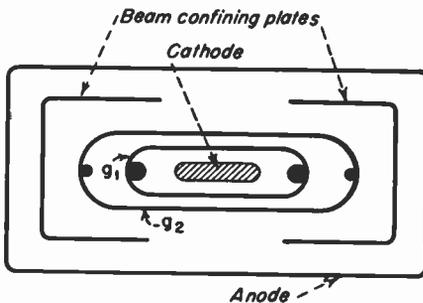


FIG. 12. Cross section of beam-power tube for low-voltage operation. Proper spacing between edges of beam-confining plates is important to performance.

a signal equal to 95 per cent of the grid-bias voltage, tubes giving best power-output efficiency and lowest distortion are taken as the prototypes. In most cases a power-output efficiency ($100P_o/i_b e_s$) of 50 per cent can be obtained with a total (rms) distortion not greater than 5 or 6 per cent.

In most a-c-d-c radio sets, the screen-grid and plate voltage are taken from the plus end of the power supply, the screen-grid connection being made directly while the plate supply loses some direct voltage due to drop in the output transformer. Under such conditions, with screen-grid voltage somewhat higher than plate voltage, the distortion is apt to be considerably greater than when the plate voltage is equal to or greater than the screen voltage. For this reason the final design of tubes for a-c-d-c sets should be made partially in view of distortion measurements made at, for example, 105 volts at the plate and 110 volts on the screen grid.

34. Converter Tubes. A converter tube is an electron tube which operates as a converter or first detector in a superheterodyne receiver and, distinct from a mixer, supplies its own local oscillator signal. Some tubes employ two electron streams so that the tube is actually an oscillator and mixer combined; each part may be considered as a triode, pentode, etc., and will not require special mention here.

The pentagrid converter employs a single electron stream, the plate current being varied in magnitude at oscillator frequency by the signal on g_1 , and by the r-f signal on g_4 . The intelligence contained in the r-f signal is then obtained as a modulation of the i-f signal whose frequency may be the sum of or difference between the oscillator and r-f carrier frequencies.

In the usual pentagrid converter there are a cathode, five grids, and an anode. The second grid frequently comprises a pair of heavy wires disposed between the side

confining plates, because of their negative charge, further reduce the potential in the retarding region to a value that is considerably less than that under a charge-free condition. Now if the edges of the beam-confining plates are too close to each other, the beam is too closely confined and the charge density may become so great that many electrons are turned back toward the screen. The result of such a condition is low power efficiency and high distortion.

In practice, the design of the beam-confining plates is ordinarily settled by making samples with varying spacing between the edges. The tubes are tested for power output and distortion so that, with proper choice of load and

rods supporting g_1 and g_2 , and is conventionally called a *hairpin grid*. Usually g_2 and g_3 are tied together internally, as both are used as screens, while g_4 is a variable-pitch control grid upon which the r-f signal is introduced. In operation, g_1 and g_2 are used in the same way as the grid and plate of an ordinary triode oscillator with g_1 self-biased.

The conversion transconductance g_c of a converter tube is the quotient of the alternating plate current component of intermediate frequency by the alternating voltage at the signal grid under a standard condition of oscillation. The characteristic is ordinarily measured with a 60-cycle signal frequency, and since this measurement is not commonly given in textbooks, it is described here. The circuit is shown in Fig. 13.

It will be recognized that if the oscillator and signal frequencies are equal, the difference frequency in the plate circuit will be zero, and that the plate current will be greater when the two signals are in phase than when they are applied 180 deg out of

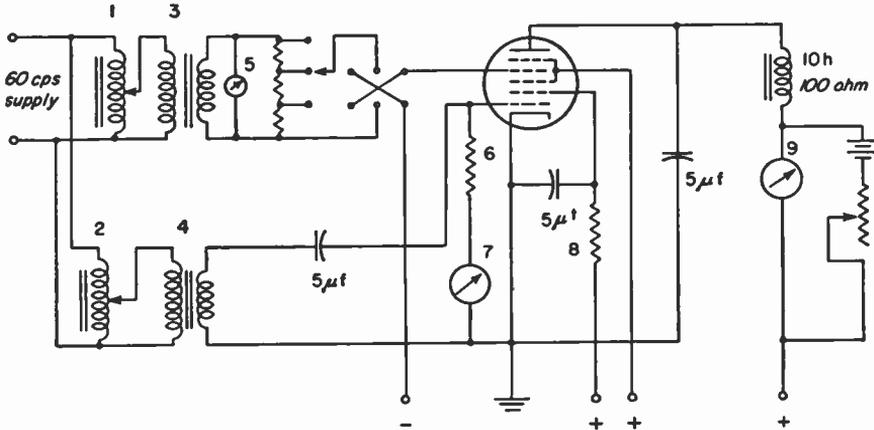


FIG. 13. Circuit for measuring conversion transconductance of pentagrid converter tubes. (1, 2) Variacs. (3) Step-down transformer, 5-volt secondary. (4) Step-down transformer, 25-volt secondary. (5) A-c voltmeter, 5 volts full scale. (6) Resistor of value specified for tube. (7) D-c milliammeter, 1.0 ma full scale. Adjust current by (2) to rated value of tube under test. (8) Dropping resistor; use when specified. (9) Laboratory-type multirange milliammeter.

phase. Therefore, the difference in direct plate current which may be measured when the phase relation of the two signals is reversed, divided by twice the peak value of alternating voltage applied to g_4 , is the conversion transconductance g_c .

Since it is neither convenient nor desirable to have the tube operating as an oscillator to provide both signals for such a measurement, the signal is taken from the 60-cycle supply as indicated in the figure. The magnitude of the signal to g_1 is commonly adjusted for each tube so that the grid current will have the value rated for the oscillation condition at the specified value of the grid resistor 6 in Fig. 13.

A bucking circuit is placed in the plate lead so that the plate current difference Δi_b , obtained when the reversing switch is thrown, may be read accurately; the current in the meter is first adjusted to zero when the signals are out of phase. The conversion transconductance will, with such a test, be given by the relation

$$g_c = \Delta i_b / 2E_{g_4} \sqrt{2}$$

where E_{g_4} is the rms voltage applied to g_4 . With most tubes it is convenient to set E_{g_4} at 0.354 volt so that g_c in micromhos will be numerically equal to Δi_b when the latter is read in microamperes. In any event, E_{g_4} should be held to a low value, and

the wave form of the line voltage should be reasonably free of distortion. The usual precautions with respect to constancy of direct supply voltages and the adequacy of by-passing are to be observed.

A most important design feature of the pentagrid converter is the variable-pitch control grid, g_4 . This grid operates at a negative voltage, the magnitude of which is controlled by the a-v-c circuit of the receiver. A curve showing g_c plotted against e_{c4} should, therefore, resemble in a general way the shape of the g_m - e_c curve of the pentode amplifier used in the same equipment. If the winding of g_4 were constant, the g_4 -plate transconductance, hence the conversion transconductance, would be sharply peaked at a value of e_{c4} which would depend upon the pitch and upon the dimensions and voltage of g_4 . Furthermore, if the grid were made with a number of sections having different pitches, the curve would be full of humps, each hump corresponding to a certain pitch. The reason why this would be so is obvious in the light of what has been said of the suppressor-grid voltage-plate current relations in a pentode. The conditions in the region between g_3 and g_4 of a pentagrid converter are similar in that a negative grid has a positive electrode each side of it, so that control consists in turning back a portion of the electrons that approach it.

In view of the fact that each section produces a peak in g_c at a different value of e_{c4} , if a smooth curve is to be obtained the pitch of g_4 must be varied almost continuously throughout its length. In practice, the most widely spaced turns are in the middle of the grid, the spacing between turns being decreased with a reasonable degree of uniformity from the middle to the ends. The turns at the ends are spaced to result in maximum g_c being obtained at a value of e_{c4} about 1 volt less negative than the bias value designed for minimum signal; *i.e.*, at $e_{c4} = -2.0$ volts in most receiving tubes.

Sometimes control grids g_4 are found having several turns of constant pitch at each end. This practice can be overdone, and it results in a very steep slope in the curve of g_c against e_{c4} in the region of -3 volts with considerable spread in the value under standard test conditions, usually $e_{c4} = -3.0$.

A further result of improper design of g_4 is an abnormally large shift in g_c during the first 50 or 100 hr of operation. In the section on triodes, attention was called to the manner in which the internal correction voltage ϵ influences i_b and g_m , and to the fact that, by proper aging, ϵ can be reduced to a value low enough so that, when the tube is designed for such a value, the deterioration of the characteristics on life is considerably reduced. In a pentagrid converter the slump in g_c during early life is due almost exclusively to change in the contact potential of g_4 . Because the grid is so far from the cathode, it does not respond readily to the 30-sec hot shot described for triodes, so that recourse must be had to such help as can be obtained by spraying the grid and by avoiding designs that result in too steep a slope in the g_c - e_{c4} curve in the region of 3 to 5 volts bias.

35. Interelectrode Capacitance. The interelectrode capacitances of electron tubes may be computed to the extent that the end or edge effects may be ignored. The shunt capacitances added by the stem, base, etc., are usually taken by measurement of the complete tube, then of the base and stem with tube electrodes removed. For adjustment of capacitances due to stem and base, a judicious choice of basing arrangement will in many cases be indicated, as well as arrangement and spacing of lead wires in the stem.

The basic formulas, as in the case for perveance, are given first for diodes, then expanded for tubes having grids. The dielectric constant is unity for vacuum, hence does not appear in the formulas.

For parallel planes,

$$C = 0.0885A/x \quad \mu\mu\text{f} \quad (51)$$

where A = area, sq cm
 x = spacing, cm

For coaxial cylinders,

$$C = 0.2416L/\log_{10}(r/r_0) \quad \mu\mu\text{f} \quad (52)$$

where L = length, cm

r and r_0 = anode and cathode radii, respectively

Approximate compensation for edge effects can be made in Eq. (51) by increasing A by one-half the product of periphery and spacing and in Eq. (52) by increasing L by $(r - r_0)$.

For Figs. 7a and b, Eq. (52) may be used with r replaced by r_x and r_y , respectively.

For tubes having grids, the cold capacitances are readily expressed in terms of the capacitance between conducting sheets located at the grids and the capacitances between the actual grids and the conducting sheets.

In the *triode*, the conducting surface that replaces the grid, here called the *grid surface*, has capacitance C_1 between it and the cathode, and capacitance between the grid surface and anode is C_2 . The grid and grid surface are coupled by C_g , related to C_2 and μ by

$$\mu = C_g/C_2 \tag{53}$$

where μ is taken to be μ_{pgk} .

C_1 and C_2 are calculated from Eq. (51) or (52) depending upon shape, x being the distance between parallel planes, r_0 and r being the inner and outer radii of coaxial cylinders.

The equivalent network gives, for the cold triode, the input capacitance

$$C_{in} = \frac{\mu(C_1 + C_2)}{\mu + 1 + (C_1/C_2)} \tag{54}$$

the output capacitance,

$$C_{out} = \frac{\mu C_2 + C_1}{\mu + 1 + (C_1/C_2)} \tag{55}$$

and the grid-plate capacitance

$$C_{gp} = \mu C_2/(\mu + 1) \tag{56}$$

For the *tetrode*, C_2 is the capacitance between g_1 and g_2 surfaces, C_3 is the capacitance between the g_2 surface and plate, and the capacitance between the g_2 surface and actual g_2 is $C_{g_2} = \mu_2 C_3$ where μ_2 is the amplification factor of the tube when g_1 is omitted. The cold capacitances with g_2 grounded through a by-pass having negligible reactance are

$$C_{in} = \frac{\mu_1(C_1 + C_x)}{\mu_1 + 1 + C_1/C_x} \tag{57}$$

where $C_x = \frac{C_3(\mu_2 + 1)}{1 + C_3(\mu_2 + 1)/C_2}$

μ_1 = triode mu, plate at screen-grid surface

μ_2 = triode mu, g_1 omitted

$$C_{out} = \frac{\mu_2 C_3 + C_v}{\mu_2 + 1 + (C_v/C_3)} \tag{58}$$

where $C_v = \frac{\mu_1 C_2}{\mu_1 + \mu_1(C_2/C_1) + 1}$

$$C_{g1-p} = \frac{\mu_1 C_2}{1 + \mu_1(C_2 + C_x)/C_3} \tag{59}$$

where $C_x = \frac{C_2 + \mu_2 C_1 C_3}{C_1 + \mu_2 C_3}$

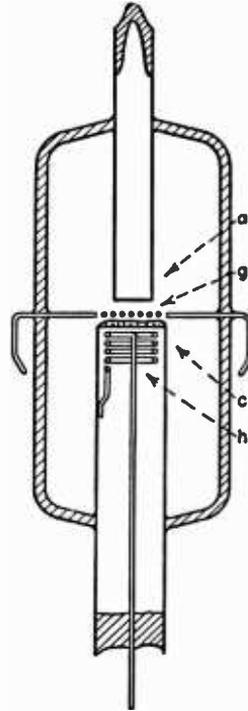


FIG. 14. Microwave triode. The principal parts are anode a , grid g , and cathode c . The cathode is coated on the end facing the grid and is heated by coiled heater h .

For *pentodes*, the corresponding capacitances are taken by replacing g_3 with the grid surface which is coupled to g_3 through the capacitance $\mu_4 C_5$. However, the resultant value of grid-plate capacitance is smaller than the actual end effects even with practicable shielding. For most purposes the input capacitance may be taken as that of the triode with plate at g_2 , and the output capacitance as that of a triode with the cathode at g_2 , since g_2 will be by-passed to ground.

36. Effect of Space Charge. The presence of electrons in the space between the electrodes increases the capacitance, since the field at the anode is greater than when the interelectrode space contains no electrons. For parallel plates in which one electrode is a cathode and the other a plate to which a space-charge-limited current flows, the capacitance is $\frac{1}{3}$ greater than in the charge-free condition given by Eq. (51). This, however, is not to be taken as meaning that the susceptance increases by the same factor; in fact, the susceptance is lower in the space-charge-limited diode than between the charge-free electrodes except at extremely high frequencies. Thus, for the cold tube, the admittance y for a diode having plane electrodes is

$$y = jb_0 = j\omega C \quad (60)$$

while for the space-charge-limited diode,

$$y = g_0 + j(\frac{2}{3}\omega C) \quad (61)$$

where $g_0 = 3i/2e_p$, the zero-frequency conductance

ω = angular frequency

C = cold capacitance

In tubes having grids the characteristics are all modified by space charge, especially in the cathode-grid region. For example, Eq. (61) applies to the admittance of the equivalent diode-anode of a tube having plane electrodes, so that an approximation of the altered transadmittance Y_{po} may be written

$$Y_{po} \simeq g_m [\cos (1\frac{1}{5}C_1 + C_2) - j\omega \sin (1\frac{1}{5}C_1 + C_2)] \quad (62)$$

where C_1 is the cold capacitance between cathode and grid plane and C_2 is that between grid plane and plate. It is to be assumed that $e_p \gg e'$, the equivalent diode voltage, so that space-charge effects beyond the grid may be disregarded for a first-order approximation. A more accurate expression for the transadmittance is obtained from the equations that include the effects of transit time, shown in subsequent paragraphs.

37. Transit-time Effects. At frequencies where the time of flight of an electron between a pair of electrodes is an appreciable part of the time of one cycle of the alternating voltage between the electrodes, the characteristics are no longer described by the simple relations that have been given up to this point.

One visualizes an electron as being emitted from the cathode of an electron tube at a relatively low velocity, then being accelerated by the field of another electrode to a relatively high velocity. Regardless of the path taken or of the velocity it may have had at some previous position in flight, the kinetic energy possessed at any instant is the product of its charge by the *net* change in potential. The velocity in the direction of travel will be

$$v = (2Ve/m)^{1/2} \quad (63)$$

which, in practical units, is

$$v = 5.93 \times 10^7 V^{1/2} \quad \text{cm/sec} \quad (64)$$

where V is the potential, e and m being, respectively, the charge and mass of an electron.

When the electron passes from a plane cathode to a plane parallel anode of a diode and the current is space-charge-limited, transit time will be

$$T = 3x/v \quad \text{sec} \quad (65)$$

where x is the distance between the electrodes and v is the final velocity given by Eq. (64). Where the current is not space-charge-limited, the transit time is lowered,

the value under the condition of zero space charge being two-thirds that given by Eq. (65).

The *transit angle* θ of the region between two electrodes of a tube is the product of the transit time T of an electron through the region and the angular frequency ω of the alternating voltage between the electrodes. For a diode,

$$\theta = \omega T \quad \text{radians} \tag{66}$$

The *impedance* z of the plane diode when the electron current is space-charge-limited is given by Llewellyn¹ in the form

$$z = (12r_0/\theta_1^4)[2(1 - \cos \theta_1) - \theta_1 \sin \theta_1] - j(12r_0/\theta_1^4)[(\theta_1^2/6) + \theta_1(1 + \cos \theta_1) - 2 \sin \theta_1] \tag{67}$$

where $r_0 = de_p/di = \frac{3}{5}(e_p/i)$ at very low frequency.

It is generally more useful to take the reciprocal of Eq. (67) so that the admittance of the gap is had. Thus if Eq. (67) is abbreviated in the form

$$z_{11} = r + jx_1 \tag{68}$$

the reciprocal or admittance will be

$$y = \frac{1}{z} = \frac{r^2}{r^2 + x^2} + j \frac{x^2}{r^2 + x^2} = g + jb \tag{69}$$

It is noted that the zero-frequency resistance r_0 is a coefficient in both the resistive and reactive parts of Eq. (68); similarly, the zero-frequency conductance g_0 is part of g and b of Eq. (69). From Table 4 giving values of g/g_0 and b/g_0 , the admittance at higher transit angles may be found for values of g_0 and θ_1 . Thus, if $\theta_1 = \pi/2$, the admittance $y = 0.9418g_0 + j0.4757g_0$.

It should be noted that, in Table 4, the value of b/g_0 for θ_1 is zero. This is not, of course, to be taken as meaning that the susceptance can be taken as zero for extremely low transit angles. As the transit angle approaches zero, the admittance approaches asymptotically the value

$$y_{(\theta \rightarrow 0)} = g_0 + j(\frac{3}{5}\omega C) \tag{70}$$

where C is the cold capacitance.

In Table 4 it will be observed that g/g_0 is negative for values of θ_1 between $\theta_1 = 2\pi$ and about 9.0 radians, reaching a value of -0.2464 at $\theta_1 = 5\pi/2$. It is this fact that

Table 4. Values of g/g_0 and b/g_0 Space-charge-limited Diode

Radians	g/g_0	b/g_0	Radians	g/g_0	b/g_0
0	1.000	0	7.07	-0.189	2.981
0.4	0.996	0.120	7.2	-0.211	3.087
1.0	0.977	0.301	7.33	-0.228	3.196
1.4	0.954	0.423	$5\pi/2$	-0.246	3.645
$\pi/2$	0.942	0.676	8.4	-0.168	4.096
1.8	0.923	0.547	8.64	-0.106	4.273
$3\pi/4$	0.867	0.723	9.0	-0.004	4.499
2.8	0.811	0.867	3π	+0.135	4.709
π	0.760	0.982	10.0	+0.264	4.906
3.6	0.681	1.143	10.21	+0.290	4.964
4.0	0.603	1.291	10.47	+0.306	5.033
$3\pi/2$	0.439	1.579	11.0	+0.287	5.177
5.2	0.310	1.802	11.52	+0.215	5.351
5.6	0.197	2.005	11.78	+0.165	5.454
2π	0.000	2.409	12.04	+0.111	5.570
6.8	-0.134	2.773	4π	0.000	5.839

¹ Llewellyn, F. B., and L. C. Peterson, Vacuum Tube Networks, *Proc. IRE*, **32**, 144-166, March, 1944.

makes oscillation possible in the Llewellyn diode, for if the circuit is adjusted to have an angular frequency ω and the plate voltage is adjusted by Eq. (65) so that $T_1 = 5\pi/2\omega$, oscillations will be produced, provided the shunt conductance of the external circuit is not greater than $+0.2464g_0$.

38. Transit Angle Effects in Multigrid Tubes. In triodes and in tubes having more than one grid, the transit angles in all regions affect the characteristics of the tube. A full discussion is found in the Llewellyn reference. The manner in which the characteristics of some tubes are affected will be seen in the results given below. These equations may be taken as first-order approximations when they are restricted as follows:

1. The electrodes are parallel planes, edge and lead effects ignored.
2. The cathode current is space-charge-limited, and the virtual cathode is assumed to be at the real cathode.
3. Space charge in the region beyond the control grid is assumed to be negligible.
4. The control grid draws no current from the electron stream.

The transit angles will be taken to have the following values:

Region 1. For the region between the cathode and the first electrode, the transit angle will be that given by Eqs. (65) and (66) for the space-charge-limited diode and may be expressed in the form

$$\theta_1 = \omega T_1 = \omega x_1 \times 10^{-7}/2 \sqrt{e'} \quad (71)$$

where x_1 is the cathode-grid spacing and e' is the potential of the equivalent grid plane.

Region 2. For the region between the grid and plate of a triode,

$$\theta_2 = \omega T_2 = 3.37\omega x_2 \times 10^{-8}/(\sqrt{e'} + \sqrt{e_b}) \quad (72)$$

For a tetrode,

$$\theta_2 = \omega T_2 = 3.37\omega x_2 \times 10^{-8}/(\sqrt{e'} + \sqrt{e''}) \quad (72a)$$

where x_2 is the spacing between control and screen grids and e'' is the voltage of the equivalent screen-grid plane and may be taken to have the value

$$e'' = (\mu_2 e_{c2} + e_b)/(1 + \mu_2 + (x_2/x_1)) \quad (73)$$

where x_2 is the screen grid-plate spacing and μ_2 is the amplification factor of the tube when g_1 is omitted. The error is usually not great if e'' is taken as e_{c2} , provided the value of μ_2 is high.

Region 3. For the region between g_2 and anode,

$$\theta_3 = \omega T_3 = 3.37\omega x_3 \times 10^{-8}/(\sqrt{e''} + \sqrt{e_b}) \quad (74)$$

Although it would be understood that the equivalent diode controlling voltage e' is to include the internal correction voltage ϵ , the error is not great if it is not included in the screen grid and anode potentials; thus the direct cathode current may be expressed

$$i_k = G \left[\frac{\mu_1 e_0 + e''}{\mu_1 + 1 + \frac{4}{3}(x_2/x_1)} \right]^{3/2} = G(e')^{3/2} \quad (75)$$

where G is the perveance and e_0 is the voltage applied to the control grid plus the internal correction voltage.

For triodes,

$$\begin{aligned} \mu_1 &= \text{amplification factor} \\ e'' &= \text{plate voltage } e_b \end{aligned}$$

For tetrodes,

$$\begin{aligned} \mu_1 &= \text{triode amplification factor} \\ e'' &= \text{equivalent voltage at screen grid, taken as } e_{c2} \text{ where } \mu_2 \text{ is high} \end{aligned}$$

The total conductance di_k/de' at zero frequency is taken to be

$$\begin{aligned} g_0 &= \frac{3}{2} G(e')^{1/2} \\ &= \frac{3}{2} i_k \left[\frac{\mu_1 + 1 + \frac{4}{3}(x_2/x_1)}{\mu_1 e_0 + e''} \right] \quad (76) \end{aligned}$$

When the transit angles in both regions of the triode and all three regions of the tetrode are just great enough to be significant, the grid-plate transconductance may be taken to have the following values:

For the triode,

$$g_m = \frac{g_{0\mu}}{\mu + 1 + \frac{4}{3} \frac{x_2}{x_1} \left(1 + \frac{1}{2} \frac{\theta_2}{\theta_1} \frac{\sqrt{e'} + 2\sqrt{e_b}}{\sqrt{e'} + \sqrt{e_b}} \right)} \quad (77)$$

For the tetrode,

$$g_m = \frac{i_b}{i_k} \frac{g_{0\mu_1}}{\mu_1 + 1 + \frac{4}{3} \frac{x_2}{x_1} \left(1 + \frac{1}{2} \frac{\theta_2}{\theta_1} \frac{\sqrt{e'} + 2\sqrt{e''}}{\sqrt{e'} + \sqrt{e''}} \right)} \quad (78)$$

where μ_1 is the triode μ .

The *grid-plate transmittance* Y_{gp} , of which the real part is the transconductance, may be expressed in terms of the l-f values given by Eqs. (77) and (78).

For moderately low frequencies,

$$Y_{gp} \simeq g_m (\cos \phi - j \sin \phi) \quad (79)$$

$$\theta_1 \left[\frac{11}{30} (\mu_1 + 1) \frac{7}{18} \frac{x_2}{x_1} \right] + \theta_2 \left[1 + \mu_1 + \frac{4}{3} \frac{x_2}{x_1} \left(1 + \frac{1}{8} \frac{\theta_2}{\theta_1} \frac{3\sqrt{e'} + 5\sqrt{e''}}{\sqrt{e'} + \sqrt{e''}} \right) \right]$$

where $\phi = \frac{1 + \mu_1 + \frac{4}{3} \frac{x_2}{x_1} \left(1 + \frac{1}{2} \frac{\theta_2}{\theta_1} \frac{\sqrt{e'} + 2\sqrt{e''}}{\sqrt{e'} + \sqrt{e''}} \right) + \frac{1}{3} \theta_2 \left(\frac{\sqrt{e''} + 2\sqrt{e_b}}{\sqrt{e''} + \sqrt{e_b}} \right)}{\quad} \quad (80)$

For very high values of transit angle,

$$Y_{gp} \simeq \frac{i_b}{i_k} g_0 \left(\frac{\mu_1}{1 + \mu_1 + x_2/x_1} \right) \left[\frac{\sqrt{e''} - P \sqrt{e_b}}{j(\theta_2/2)(\sqrt{e''} + \sqrt{e_b})} \right] Q \quad (81)$$

where $P = \cos \theta_2 - j \sin \theta_2$

$Q = \cos (\theta_1 + \theta_2) - j \sin (\theta_1 + \theta_2)$

For the triode, the expressions are obtained directly from Eqs. (80) and (81) by setting $e'' = e_b$ and letting θ_2 approach zero so that the second term of Eq. (80) drops out and the bracketed term of Eq. (81) becomes unity; the equivalent triode μ , μ_1 , then becomes the triode amplification factor μ .

For the pentode, the real part of the transadmittance may be taken as substantially that of a tetrode, the phase being increased by the increase in transit angle between the screen grid and plate of the pentode as compared with that in the tetrode.

TUBES EMPLOYING BEAM MODULATION

39. Klystrons. A klystron is a tube in which an electron beam becomes bunched by passing through an alternating electric field and power is extracted from the a-c component of the bunched beam. Where the same resonant system causes both the bunching and extraction of the power from the beam, the tube serves as an oscillator and is commonly called a *reflex klystron*. Tubes employing more than one resonant system coupled to a single beam may be used as oscillators, amplifiers, frequency multipliers, modulators, or detectors.

In reflex klystrons the electrons pass twice through a single pair of resonator grids, being reflected back through the gap by the retarding field of the *reflector electrode* which is held at a negative voltage. Figure 15 shows schematically the general form of such a tube with its resonant cavity.

The cavity resonator is held at a voltage e_b with respect to the cathode and the

electrons enter the gap between the r-f grids at a velocity

$$v_0 = 5.93 \times 10^7 \sqrt{e_b} \quad \text{cm/sec} \quad (82)$$

With an alternating voltage across the gap between the resonator grids, electrons are accelerated or slowed down according to the magnitude and direction of the field at the time they leave the gap. Since the velocities of the electrons as they leave the gap and enter the retarding field between the upper resonator grid and reflector vary with time, the rate at which they are returned to the gap varies also with time. Under certain conditions the beam will be bunched as the electrons reenter the gap, and if the field across the gap is now of such polarity as to slow down the returning electrons, power is delivered to the resonator.

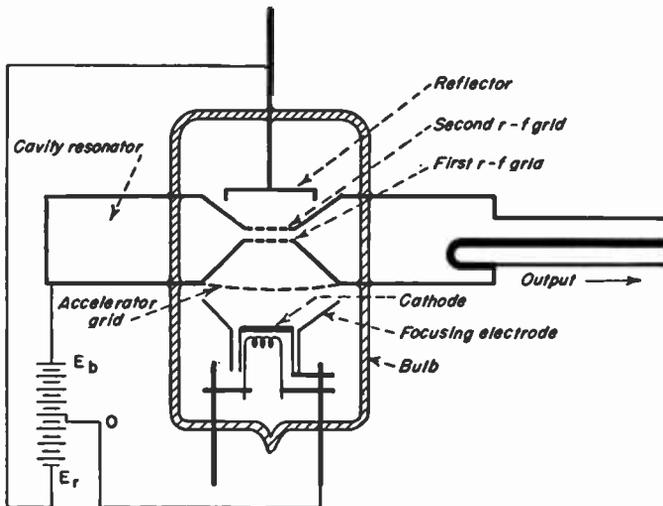


FIG. 15. Schematic drawing of reflex klystron in glass bulb with resonant cavity, loop and line attached. The space between the first and second r-f grids is the *interaction gap*. Electrons become bunched in the *drift space* located between the interaction gap and the reflector.

Those electrons making their initial transit of the gap when the instantaneous field across the gap is zero and when the gap voltage is changing from acceleration to deceleration become the centers of the bunches ultimately formed. These electrons remain in the retarding field for a time T_0 , in which

$$T_0 = \frac{4x_r}{v_0} \frac{e_b}{e_b - e_r} \quad \text{sec} \quad (83)$$

where x_r is the spacing between grid and reflector, and e_r is the potential of the reflector with respect to the cathode.

To describe the power relations of the system, it is necessary to know the "mode number" in which the tube is operating. This is expressed

$$N = fT_0 \quad (84)$$

in which f is the frequency of the resonator hence N is the ratio of the time T_0 to the period of the resonator. For maximum power delivery to the system,

$$N = n - \frac{1}{4} \quad (85)$$

in which n is an integer greater than zero.

The mode number N is a factor of the "bunching parameter" x which is written

$$x = \pi N e_1 / e_b \tag{86}$$

where e_1 is the peak instantaneous voltage across the gap.

Equation (86) is valid only when the gap transit angle $\theta_1 = \omega x_1 / V_0$ is small. For large values of θ_1 the expression is to be multiplied by the "beam-coupling coefficient" β , which is given as

$$\beta = \frac{\sin (\theta_1 / 2)}{\theta_1 / 2} \tag{87}$$

Table 5. Values of Beam-coupling Coefficient β for Certain Values of Internal Transit Angle θ_1

θ_1	β	θ_1	β
$\theta_1 \rightarrow 0$	1.000	$3\pi/4$	0.777
$\pi/4$	0.978	π	0.633
$\pi/3$	0.955	$3\pi/2$	0.300
$\pi/2$	0.905	2π	0.000
$2\pi/3$	0.830		

The general expression for the bunching parameter then becomes

$$x = \beta N \pi e_1 / e_b \tag{88}$$

The bunched electron stream passing through the gap constitutes a current which may be expressed by a Fourier series in which the coefficients are Bessel functions of the first kind. In the reflex tube the fundamental only is of use, and this may be expressed in the form

$$i = 2I_0 J_1(x) (\cos \phi - j \sin \phi) \tag{89}$$

where I_0 is the d-c value of that part of the beam which passes twice through the gap, and $J_1(x)$ is the first-order Bessel function of the bunching parameter x , Eq. (88).

Table 6. Values of the Ratio of R-f Current to Direct Beam Current for Selected Values of the Bunching Parameter $x = \pi N e_1 / e_b$

x	i/I_0	x	i/I_0
0	0	6.0	-0.55
1.0	0.8	7.0	0.0
1.9	1.2	8.0	+0.5
2.0	1.15	8.5	0.55
3.0	0.65	9.0	0.5
3.8	0.0	10.0	0.1
5.0	-0.6	11.0	-0.3
5.3	-0.65		

The value of $J_1(x)$ is maximum when $x = 1.84$; $J_1(1.84) = 0.5819$. The phase angle ϕ is given as

$$\phi = 2\pi N - 2\pi(n - \frac{1}{4}) \tag{90}$$

It is convenient to consider that the electron stream constitutes an admittance element coupled to the resonator. The current through this element is βi and the admittance is the quotient of βi by the voltage across the resonator. The admittance then will be

$$y_i = -2\beta(I_0/e_1)J_1(x)(\cos \phi - j \sin \phi) \tag{91}$$

When the phase angle ϕ is zero, the current and voltage are π radians out of phase, hence the conductance is negative.

40. Klystron Efficiency. The condition for oscillation is that the magnitude of the real part of Eq. (91) is greater than the shunt conductance of the resonator. It is assumed that, as the oscillations increase, the magnitude of the tube conductance decreases and the oscillations become stable when the total conductance across the gap is zero. The frequency will, of course, be such that the sum of the susceptances of the tube and of the resonator is zero.

The power delivered to the resonator will be the product of the rms value of the voltage and the in-phase component of βi . Eliminating the term e_1 by Eq. (88), the expression for the power is

$$P_o = (e_b I_o / \pi N) x J_1(x) \cos \phi \quad (92)$$

The phase angle ϕ is readily varied by variation of the reflector voltage. Ordinarily, tuning is accomplished by varying the resonant frequency of the cavity and adjusting the reflector voltage e_r until the output is approximately maximum. The manner in which the final adjustment to frequency may be made by variation in the reflector voltage is understood by reference to Eq. (91), wherein the susceptance is a function of ϕ which, in turn, depends upon the reflector voltage. Similarly, a signal may be applied to the reflector with the result that the output of the resonator is frequency-modulated.

The output efficiency of a reflex tube is ordinarily of the order of 1 or 2 per cent when the quotient of power delivered to the line by electrode dissipation is taken. If the total electron current from the cathode is I_b , the power dissipated is $I_b e_b$. The current I_b is clearly to be distinguished from the beam current I_o ; the latter is only that fraction of I_b which is not captured by an electrode before two complete transits of the gap have been made. For what it might be worth as a guide, one could divide Eq. (92) by the power dissipated and obtain the expression,

$$\text{Eff.} = \frac{I_o}{I_b} \times \frac{x}{\pi N} J_1(x) \quad (93)$$

as a maximum when $\phi = 0$ and the power is taken as that delivered to, not by, the resonator. Setting $I_o = I_b$, $x J_1(x) = 1.0$, and $N = 1\frac{3}{4}$, Eq. (93) is evaluated to 18 per cent, which is about ten times greater than is ordinarily obtained with tubes of good design.

Special design features have been described whereby the theoretical and actual efficiency of reflex tubes has been increased. In one modification the reflector has a surface adapted to emit secondary electrons when bombarded by primary electrons. When the reflector is held at a potential intermediate between those of resonator and cathode and when bunching is made to occur at the reflector, the reflector then emits a density-modulated electron stream which, if properly phased, delivers power to the resonator.

In ordinary reflex klystrons most of the electrons that succeed in penetrating the meshes of all grids twice are captured by the electrodes before they are able to make a further transit of the gap. Tubes have been made which are so designed that a considerable fraction of the electrons make more than two transits of the gap with the delivery of additional power. The efficiency claimed for these tubes is greater than that obtained with the ordinary tube in which the number of electrons making more than two transits of the gap is negligible.

41. Two-cavity Klystron. A generalized diagram of a two-cavity klystron is shown in Fig. 16. Here the bunching action is instituted by the alternating voltage across the input resonator grids, and a condition necessary for amplification or power generation is that the beam shall be at least partly bunched as it passes through the output resonator grids.

The signal to be amplified is introduced into the bunching cavity by any convenient means, ordinarily a loop. When the tube is used as an oscillator, feedback may be accomplished by coupling between input and output resonators. As a frequency multiplier, the tube is used in the same manner as for amplification, the input resonator

being tuned for the input signal frequency and the output resonator tuned for the desired harmonic.

The drift space in a two-cavity klystron is generally greater than twice the distance between resonator and reflector of a reflex tube. In the latter the electrons come to a stop, whereas in the two-cavity tube the electron velocity is substantially constant. The average transit time will be

$$T_0 = x_d/v_0 \quad \text{sec} \tag{94}$$

in which x_d is the length of the drift space. For maximum power delivery to the out-

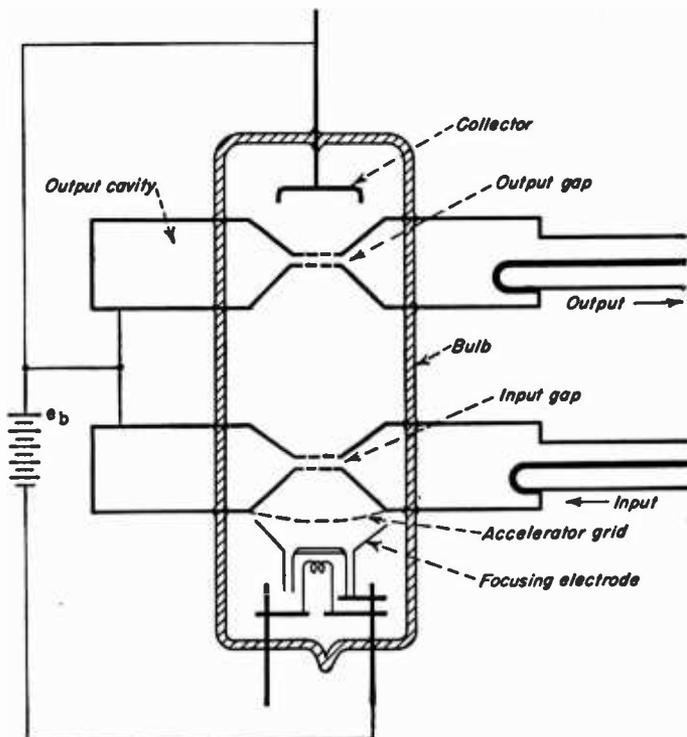


FIG. 16. Schematic drawing of two-cavity klystron. Bunching action is begun in the input gap. In the output gap the bunching of the electrons reaches an optimum value. This type of tube may be used for amplification at microwave frequencies or for frequency multiplication.

put resonator grids; where internal feedback is used, the mode numbers are the same as the reflex tube, *i.e.*, $N = fT_0(n - \frac{1}{4})$ where n is an integer greater than zero. When used as an oscillator, the tube is usually operated at $N = \frac{3}{4}$.

The two-cavity klystron is important as a frequency multiplier. For a constant level of power output, it is required that the magnitude of the signal voltage across the input gap be considerably increased as the frequency multiplication factor M is increased. The current through the output resonator gap is then of the form given in Eq. (89) except that the M th Bessel function is taken. When $M = 1$, the optimum value of x is 1.84, but for higher values of M , $J_M(x)$ will be maximum for lower values

of z . The beam-coupling coefficient β will, of course, be taken for the transit angle at the output frequency in the output resonator gap.

If the input frequency and power are held fixed, the output power drops rapidly as M is increased, since the values of the Bessel functions of the bunching parameter, the beam-coupling coefficient, and the shunt resistance of the unloaded output resonator all decrease. However, if the output frequency is constant and the signal frequency is decreased, satisfactory performance is usually obtained with values of M at least as high as 10.

Since the two-cavity klystron may be operated with a very low value of N and there is practically no space charge in the drift region, the power-conversion efficiency is much higher than in the reflex tube. For $M = 1$, $N = \frac{3}{4}$, and discounting loss of beam current by grid capture, the maximum theoretical efficiency would be about 58 per cent. For higher values of M with constant output frequency, the output power of the M th harmonic varies approximately as $M^{-\frac{1}{2}}$.

Klystrons have been made with more than two interaction gaps or pairs of grids. Bunching is started by the low-level signal applied to the first resonator, and amplification results. The amplified signal may then be applied to the electron beam with the result that the bunching is intensified at the final output resonator.

42. Magnetrons. The word "magnetron" has been used for many years to describe a class of electron tubes in which a magnetic field is used to direct electrons in a curved path through an electric field in such a manner that oscillations are induced into a resonant system. During the Second World War the development of the magnetron was accelerated to provide a microwave signal at high power for use in radar and pulse-operated communication systems. The result was tubes that operated at and below a wavelength of 10 cm with peak-power outputs ranging from about 100 watts for low-power communication equipment to several megawatts for high-power radar.

The tubes designed for high-power pulsed operation at microwave frequencies usually make use of a cylindrical, coated cathode and an anode or body of oxygen-free, high-conductivity copper with the cavity resonators milled or hobbled out of a solid block or formed by hard-soldering vanes around the inside of a cylinder.

The tube is operated with the body grounded and the cathode at a negative voltage which is ordinarily pulsed so that space current flows only a relatively short portion of the time. The pulses are ordinarily of one or a few microseconds duration.

In addition to the electric field produced by the anode voltage, a magnetic field is applied parallel to the axis of the tube. The magnetic field exerts a force on the moving electrons which is perpendicular both to the direction of the magnetic field and to the direction of the motion of the electron.¹

At any point in the space between cathode and anode, the force on an electron due to the electric field is

$$f(e) = -eF \quad (95)$$

where e is the charge on the electron and F is the field, and the velocity is

$$v = \sqrt{2Ve/m} \quad (96)$$

where m = mass of electron

V = potential at point

The force on the electron due to the magnetic field of strength B is

$$f(m) = Bev \quad (97)$$

The quantities F and B are vectors; we consider here the components which are mutually perpendicular. Resolving the centripetal force of the electric field and the centrifugal force of the magnetic field, the radius of curvature r of the electron path

¹ The discussion in the succeeding paragraphs is intended to give only a rough idea of the way a magnetron operates as an oscillation generator. For a reasonably complete discussion of the theory and a description of practical tubes developed during the war, the reader is referred to the reports of the M.I.T. Radiation Laboratory and Fisk, Hagstrum, and Hartman, *The Magnetron as a Generator of Centimeter Waves*, *Bell System Tech. J.*, 25 (2), 167-348, April, 1946.

at the point is given by

$$r = mv/eB \tag{98}$$

which, in practical units, becomes

$$r = 3.37V^{1/2}/B \quad \text{cm} \tag{99}$$

In view of Eq. (99), it is evident that the magnetic field and the voltage of the anode can be so adjusted that the electrons just barely graze the anode before they are turned back to the cathode. Under these conditions the anode voltage and magnetic field are in the relation

$$\frac{e_a}{B^2} = \frac{e}{m} \cdot \frac{r_a^2}{8} \left[1 - \left(\frac{r_c}{r_a} \right)^2 \right]^2 \tag{100}$$

When converted to practical units, Eq. (100) is expressed

$$\frac{e_a}{B^2} = 0.022 r_a^2 \left[1 - \left(\frac{r_c}{r_a} \right)^2 \right]^2 \tag{101}$$

where e_a = anode voltage
 B = field, gausses

r_a and r_c = anode and cathode radii, cm

During the excursion of an electron into the interelectrode space, it will acquire an amount of energy eV , where V is the potential at its closest approach to the anode. This is exactly the amount which will be expended in getting back to the cathode. Therefore, when the electron gets back to the cathode, its velocity is exactly that which it had initially.

Now suppose that energy can be added to the electron by additional means. Referring to Fig. 17, let the path be that indicated as a , and assume that as the electron passes the gap there is a potential difference across the gap which increases the velocity or energy of the electron. The points of particular interest here are that the electron takes energy from the source that supplied the potential difference, and that this energy is in excess of the amount required to get back to the cathode. The result is that the energy picked up in passing through the field across the gap will be dissipated in the form of heat when the electron strikes the cathode.

On the other hand, suppose that an electron passes a gap when the voltage is such as to decrease the velocity of the electron. The decrease in velocity represents a transfer of energy from the electron to the cavity, and the electron no longer has sufficient kinetic energy to penetrate the field all the way back to the cathode. When the magnitudes of the anode voltage and the magnetic field are uniquely adjusted to the physical dimensions of the system, the path b of those electrons which give up energy to the system in their first passage by the gap is such that, in subsequent passages of the gaps, they transfer additional power to the cavities. The fact that, for every electron that absorbs energy from the resonant system once, there is another electron that delivers energy more than once is the reason why the system continues to oscillate. As shown in the figure, the paths of the electron which deliver energy to the system terminate on the metallic conductors.

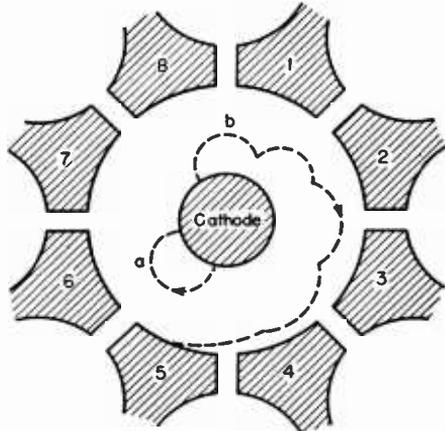


FIG. 17. Typical electron paths in an eight-cavity magnetron. Electron which follows path a absorbs power from the system once, then returns to the cathode. Electron following path b delivers energy every time it passes a gap, finally arriving at the anode.

43. Magnetron Modes. A magnetron may oscillate in a number of modes unless special features are built into the tube which prohibit or decrease the chance that undesirable modes shall be present. These modes are to be distinguished from the modes of a klystron or of a simple cavity resonator. The term is used to describe the phase relation between the individual cavities of the tube. For example, if at a given instant the voltage across all gaps is maximum and adjacent cavities are 180 electrical degrees or π radians out of phase with each other, the tube is said to be operating in the π mode. In a tube having n cavities, oscillation may take place in any mode N such that

$$N = \phi n / 2\pi \quad \text{radians} \quad (102)$$

where n is the number of cavities and ϕ is the phase angle. N may be any integer equal to or smaller than $n/2$.

The tube shown in Fig. 17 has eight cavities; therefore, it may operate in any of four modes as follows:

N	ϕ	N	ϕ
4	π	2	$\pi/2$
3	$3\pi/4$	1	$\pi/4$

The power output obtained at the different modes is not the same and the wavelengths of the different modes are unequal. In many tubes the π mode is the most efficient, but the tube is apt to be erratic in operation, jumping back and forth between the π mode and that nearest to it where the change in wavelength is least.

To inhibit this effect, called *moding*, the impedance between alternate cavities is made low by *strapping* so that the wavelength difference for adjacent modes is increased.

Strapping consists in hard-soldering a pair of copper or silver rings at each end of the anode just behind the inner edge of the gaps. For example, in Fig. 17 one ring at each end will be connected at 1, 3,

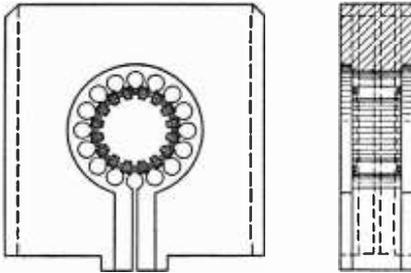


FIG. 18. A 3-cm magnetron anode.

5, and 7; the other will be connected at 2, 4, 6, and 8. These rings are usually not complete circles; e.g. the ring connecting the odd-numbered members in Fig. 17 may have a break between 1 and 7; at the opposite end of the anode it may have a break between 3 and 5. These breaks in the rings are made for the purpose of preventing the occurrence of modes that could absorb power or result in unstable operation which might be as serious as a complete lack of strapping. Figure 18 shows a 3-cm magnetron anode shown for the purpose of illustrating one type of strapping and to show the wave-guide slot at the back of one cavity through which the power is coupled to the wave guide.

44. Magnetron Adjustment. While magnetrons of the type described are intended for operation over a narrow band of frequency, a certain amount of tuning is possible through adjustment of the load which is coupled by means of a loop above one of the cavities or through the slot. In some tubes additional operating latitude is obtained by means of a movable ring, sometimes provided with teeth, which increases or decreases the gap capacitance as it is moved toward or away from the ends of the vanes. A modification comprises the use of plungers which vary the inductance of the cavities by effectively changing their volume.

It has been indicated that, if an electron that gives up energy to a cavity in its first

passage past the gap is to be properly phased to deliver energy in subsequent passages across gaps, the voltage of the anode and the magnetic field strength must be adjusted to the physical dimensions and the mode in which the tube is operating. This adjustment will result not only in causing the electrons to pass close to a gap in their first trip toward the anode; in subsequent trips they must also pass a gap at an instant when the field across the gap is the same as it was when these electrons passed the first gap. These conditions are fulfilled when the anode voltage E_0 and magnetic field B_0 are related to the other variables as follows:

$$E_0 = \left(\frac{r_a}{\lambda N} \right)^2 \times 10^7 \quad \text{volts} \quad (103)$$

$$B_0 = \frac{2.14 \times 10^4}{[1 - (r_c/r_a)^2] \lambda N} \quad \text{gausses} \quad (104)$$

where λ is the free-space wavelength of the oscillations, r_c and r_a are the cathode and anode radii, respectively. It will be noted that Eqs. (103) and (104) together give Eq. (101). These equations are based on small-signal theory and are not to be taken as holding rigorously in the practical case; they do, however, serve as reasonably good guides.

It is frequently undesirable or impossible to set the anode voltage and magnetic field at the values specified by Eqs. (103) and (104). While the condition for grazing incidence for a given tube will always be $e_b/B^2 = \text{constant}$, some deviation from this condition is tolerable while still maintaining oscillation in a given mode. When the actual field B does not differ greatly from B_0 , the proper anode voltage e_b will be given approximately by the relation

$$e_b = E_0 \left(\frac{2B}{B_0} - 1 \right) \quad (105)$$

45. Interdigital Magnetron. In this tube the anode comprises two sets of intermeshed fingers, the two sets being connected to opposite poles of a single cavity resonator. The cylindrical cathode is placed at the axis of the anode, the magnetic field being applied axially as with the multicavity tubes.

With a disk-seal type of tube operated with an external cavity, a certain amount of tuning may be achieved by adjustment of the effective dimensions of the cavity.

The relation of the oscillation frequency to the physical and electrical dimensions is the same as for the cavity magnetron. However, the fact that only one cavity is used rules out modes other than those in which adjacent fingers of the tube are either 0 or 180° out of phase. The simplest mode corresponds to a TM_{01} mode in a cylindrical cavity loaded by the finger system. There are, however, other useful modes corresponding to TM_{11} , TM_{21} , etc., modes in a cylindrical cavity. These modes may also be considered as modes of oscillation of a reentrant transmission line or guide, which are visualized as standing waves occurring around the zigzag path between the fingers.

Certain control features are incorporated into the design of a tube to strengthen oscillations in a particular mode. For example, the TM_{01} mode couples strongly to the cathode. Therefore, the power may be coupled out through the cathode lead, or a $\lambda/4$ choke may be placed in the cathode lead and the power coupled out of the cavity by means of a loop.

Tubes intended for oscillation in one of the reentrant line modes usually employ phase-reversing anodes which aid in presenting quasi π -mode fields to the electron interaction space while at the same time tending to suppress the TM_{01} mode. The principle of the phase-reversing anode will be explained.

When a tube is operating in a line mode, the field about the cathode is of the form $\sin \theta \times \sin (n\theta/2)$ for a single standing wave (the TM_{11} mode), $\sin 2\theta \times \sin (n\theta/2)$ for two standing waves, etc. Thus in the TM_{11} mode there will be two diametrically opposite points at which the voltage between the fingers of a symmetrical tube is

zero; similarly, there are four equally spaced points of zero field for the TM_{21} mode, etc. These points, in a symmetrical tube—*i.e.*, one in which the fingers are all of the same dimensions—might easily be pushed about by the loop into positions unfavorable to the transfer of power from the cavity. Furthermore, a phase slip of 180° would be required at each point of zero field in order that the phase of the electrons be proper to the delivery of power to the anode system. By replacing two pairs of fingers on diametrically opposite sides of the anode with fingers of double width (plus an additional width equal to the finger spacing) the voltage nodes are locked in place and the phase reversal of the field occurs without slip in the phasing of the electrons.

In Fig. 19, *a* is a cross section of an 18-tooth symmetrical anode intended for operation in the TM_{01} mode, while *b* shows the wide-finger modification or phase-reversing

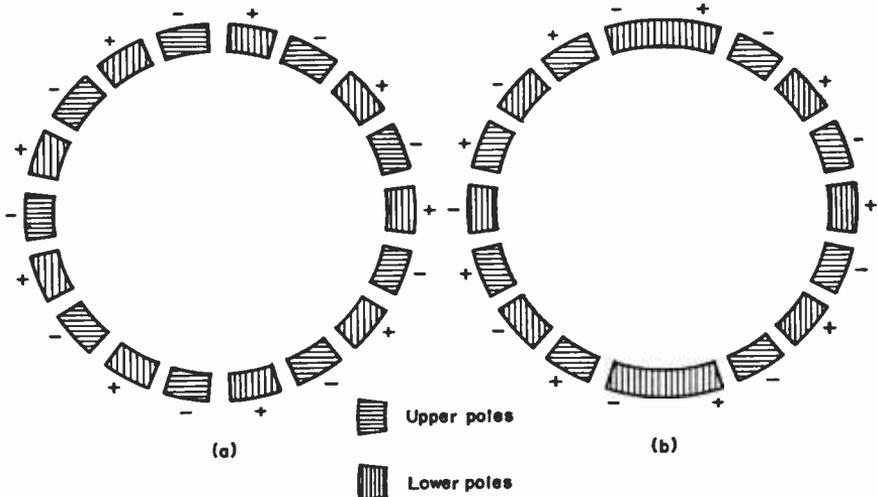


FIG. 19. Phase relations in the interdigital magnetron: (a) symmetrical anode in the TM_{01} mode; (b) phase-reversing anode in the TM_{11} mode.

anode intended for operation in the TM_{11} mode. It will be noted that the plus and minus signs of *a* and *b* alternate in the same manner; *i.e.*, the signs are reversed at 10° intervals. It follows, despite the fact that anode *b* has actually only 16 fingers, it is treated as an 18-tooth anode in the equations. In general, the number n , as understood in Eq. (102), is the number of small fingers plus twice the number of wide fingers. Also, N as used in Eq. (104) is $n/2$.

It will be noted in Fig. 19*b* that both of the wide fingers are attached to the same anode half. This will be true whenever $n/2$ is an odd number; when $n/2$ is even and the anode has two wide fingers, the latter will be on opposite halves of the anode.

Two advantages of the line-mode of operation are that the modes are widely separated, as compared with the case for multicavity tubes, and the electrodes can be relatively large for short wavelengths. In general, the free-space wavelength is of the order of one-half the length of the zigzag path between the fingers for operation in the TM_{11} mode, one-fourth the distance for the TM_{21} mode, etc. Efficiencies in the TM_{01} and the line modes are about the same, being of the order of 50 per cent.

BEAM TRAVELING-WAVE TUBES

A beam traveling-wave tube consists essentially of an evacuated enclosure containing an electron gun, a helix to which is attached input- and output-signal terminations, and an electron collector. Its principal application is in broad-band amplification

at microwave frequencies. The tube and battery supplies are shown schematically in Fig. 20. An electromagnet, not shown, is generally employed with its field along the axis of the helix to prevent spreading of the electron beam.

As shown in the figure, the input signal is applied at the gun end of the helix. In traveling along the helix the signal is, under proper conditions, amplified by interaction with the electron beam so that the power transferred to the output waveguide may be 100 or more times greater than that introduced at the input.

46. Traveling-wave Tube Principles. When a signal of intensity E_0 is applied at the input, it travels along the helical wire at a velocity which may be taken as substantially equal to that of light. Measured along the axis of the helix, however, the velocity of the wave will be considerably reduced. For the ordinary case, the axial velocity may be taken as

$$v \approx cs / (s^2 + 4\pi^2 a^2)^{1/2} \quad \text{cm/sec} \tag{106}$$

where s = distance between turns of helix

a = radius of helix

c = velocity of light = 3×10^{10} cm per sec

Electrons supplied by the gun are made to pass axially through the helix after which they are collected by the electrode shown at the end of the tube. The average velocity

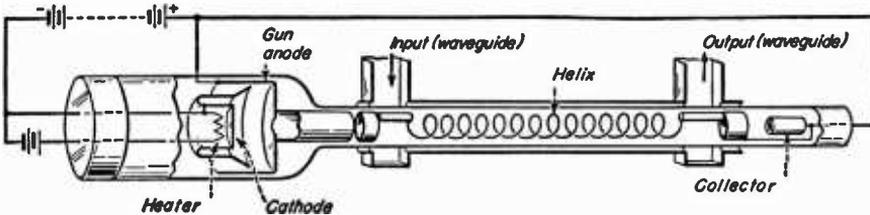


FIG. 20. Cross section of beam traveling-wave tube showing arrangement of the elements. An electron beam from cathode to collector passes through the helix and interacts with a wave traveling along the helix to produce amplification. An attenuator (not shown) is placed along the helix to choke out waves propagated backward from the output end of the tube. (Bell Laboratories Record.)

v_0 of the electrons will depend upon V_0 , the potential of the helix with respect to cathode:

$$v_0 = 5.93 \times 10^7 \sqrt{V_0} \quad \text{cm/sec} \tag{107}$$

In operation, V_0 is set so that v_0 and v are approximately equal.

In the analysis of the propagation of a wave in the helix with the electron beam current flowing, it has been shown that the signal is split into four waves each having its own direction, velocity, and attenuation. The attenuation of one of these waves is negative; *i.e.*, its power increases. With the average velocity of the electrons somewhat greater than the velocity of the increasing wave, the electrons tend to become crowded together in regions of retarding field where they give up energy to the wave. The resulting gain G may be expressed by the relation

$$G = 20 \log_{10} (E_1/E_0) \quad \text{db} \tag{108}$$

where E_1/E_0 is the ratio of the magnitudes of output and input signals. For a lossless helix the gain may be expressed in the form

$$G = A + BCN \quad \text{db} \tag{109}$$

where $A = -9.54$ db and $B = 47.3$ db.

N is the length of the system in wavelengths and is expressed as

$$N = Lf/v$$

where L = axial length of helix and f = frequency.

The quantity C is called the gain parameter and has the value

$$C = p(I_0/8V_0)^{3/2} \quad (110)$$

where I_0 is the current in the collector and p is the impedance parameter having the value

$$p = (\beta/\beta_0)^{1/2}(\gamma/\beta)^{3/2}F(\gamma\alpha) \quad (111)$$

where $\beta = \omega/v$

$$\beta_0 = \omega/c$$

$$\gamma^2 = \beta^2 - \beta_0^2$$

$F(\gamma\alpha)$ = factor shown plotted in Fig. 21

In practice, the helix is ordinarily wound so that β/β_0 will have a value of about 10 to 20; for a ratio of 13, the direct voltage V_0 will be about 1,650, a convenient value.

Feedback from output to input is attenuated by winding the helix of wire having considerable loss or by the insertion of an attenuator at some place along the helix. When lossy wire is used, the attenuation should be of the order of 20 db in the absence of the beam. For the high-conductivity helix, the attenuator may consist of a coating

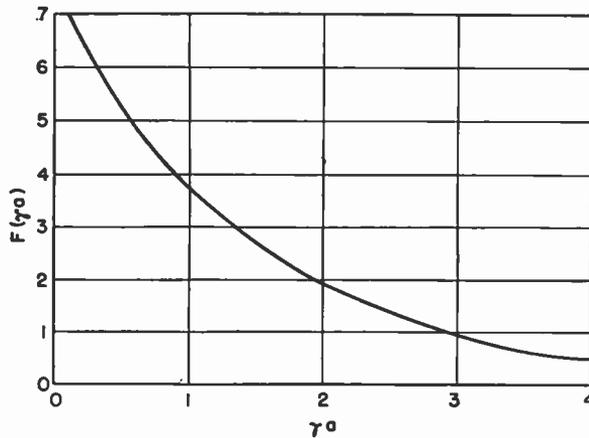


FIG. 21. Factor employed in Eq. (111) for beam traveling-wave tube.

of graphite applied to a short section of one or more of the nonconducting rods in which the helix is supported.

The reason for the relatively flat response of the traveling-wave tube is partially seen when variation in the values of the factors N and C of Eq. (109) with frequency are considered. For example, if the wavelength is doubled, there are only half as many waves between input and output, hence the number of times an electron may deliver energy to the system is halved. On the other hand, the field within the helix does not decrease toward the axis so rapidly when the wavelength is increased. Inspection of Fig. 21 shows that if $(\gamma\alpha)$ is 2.5 originally, variation in α (which is nearly proportional to variation in frequency) produces a nearly equal but opposite variation in N and C .

With good design, tubes may be made in which the noise factor is very low. Kompfner¹ reported a value of 11 db using a long, low-loss helix and low beam current, about 100 μa . Where higher beam current is used and the power gain is high, the noise factor becomes less favorable. Theory indicates that the noise factor should be best when the gain per wavelength is held to a minimum; *i.e.*, the value of C , Eq. (110), is to be low. To maintain a high value for gain and at the same time accomplish this with minimum noise, the length of the helix should be as great as practicable.

¹ Kompfner, R., The Traveling-wave Tube as Amplifier at Microwaves, *Proc. IRE*, 35 (2), 125-128, February, 1947.

CATHODE-RAY TUBES

Cathode-ray tubes ordinarily employ indirectly heated, oxide-coated cathodes. The electrodes are usually made of stainless steel or nickel. The magnitude of the electron beam is controlled by means of a grid comprising essentially a disk having a single hole. The beam may be focused on the screen by either an electrostatic lens or a magnetic lens. Similarly, the beam may be deflected on the screen by either electric fields or magnetic fields.

47. Focusing. In most cases it is easier and more economical to use electric focusing and deflection, as is done in oscillographic and low-price television equipment. Magnetic focusing produces a better spot, and magnetic deflection results in less deflection defocusing. This permits wider angle deflection, hence a shorter tube. Thus, although somewhat more expensive, all-magnetic tubes are at present preferred in high-quality television receivers.

The focusing and deflection of electrons in a cathode-ray tube depend upon the forces mentioned in connection with the magnetron. These forces are vectors, as was pointed out, and although the quantitative relations that will be given are in scalar notation, it should be understood that the reference axes must be chosen to make this convenience valid. Thus the electric force $f_{(e)}$ was given as

$$f_{(e)} = -eF$$

where e is the charge and F the field. Taking the perpendicular axes x , y , and z , the forces will be

$$\begin{aligned} f_x &= edV/dx \\ f_y &= edV/dy \\ f_z &= edV/dz \end{aligned}$$

where f_x is the force in the x direction, $-dV/dx$ is the field in the x direction, etc.

Similarly, the magnetic force was given as

$$f_{(m)} = Bev$$

where B is the magnetic field perpendicular to the path of the electron and v is the electron velocity. The forces in the three directions will be

$$\begin{aligned} f_x &= e[B_z(\partial y/\partial t) - B_y(\partial z/\partial t)] \\ f_y &= e[B_x(\partial z/\partial t) - B_z(\partial x/\partial t)] \\ f_z &= e[B_y(\partial x/\partial t) - B_x(\partial y/\partial t)] \end{aligned}$$

where the B s are the magnetic-field components in the subscript directions and the derivatives are the velocity components in the indicated directions.

While the application of these principles to the design of a satisfactory focusing system is a highly specialized problem, the deflection for small angles may readily be derived.

48. Electrostatic deflection, neglecting edge effects, requires that the deflection field be the quotient of the voltage between the plates by the distance between them. The force will be applied during the time $t = l/v$, so that the electrons emerge at an angle given by

$$\tan \phi = (l/2)(F/V_0) \tag{112}$$

where ϕ = angle of deflection
 l = effective length of deflection system
 F = average field perpendicular to axis
 V_0 = beam potential

Neglecting second-order effects, the displacement d of the spot on the screen may be written

$$d \simeq (Ll/2x)(e_d/e_0) \tag{113}$$

where L = distance from center of deflection system to screen
 x = separation between deflection plates
 e_d = voltage between deflection plates
 e_0 = voltage between deflection system and cathode

49. Magnetic Deflection Systems. Here it is assumed that the electrons enter the deflection system along the z axis. With the magnetic field perpendicular to the z axis or axis of the tube, the electrons will have a force applied in the $+y$ direction at first, then in the $+y$ and $-z$ directions as the velocity component in the y direction becomes finite. Therefore, the electrons describe a circular path in a plane perpendicular to the applied magnetic field, so that they emerge at an angle ϕ from the axis given by

$$\phi = (lB_x/V_0) \sqrt{e/2m} \quad (114)$$

where B_x is the field perpendicular to the axis, m is the mass, and e the charge. For practical purposes, the displacement of the spot on the screen will be

$$d \simeq LlB/3.37 \sqrt{V_0} \quad (115)$$

where B is the field in gauss.

50. Defocusing Effects. Both systems have inherent defects due to the fringing fields at the edges of the focusing system operating on a beam of finite diameter, the net effect being known as *deflection defocusing*. Some compensation is achieved through careful attention to the shapes of the deflection fields. In general, deflection distortion is minimum when the length of the deflection system is maximum, the limit being to a considerable degree a compromise between distortion, mechanical size, and circuit loading.

51. The Ion Trap Gun. The beam of a cathode-ray tube consists almost entirely of electrons, but there is present in addition ions of various elements in the tube, the most prevalent generally being that of oxygen, O_2^- . Where the ions strike the screen, they cause a reduction in the luminous efficiency of the phosphor. In a system employing electrostatic focusing and deflecting, the ions are focused and deflected in the same manner as are the electrons, since the deflection is independent of mass as seen by Eq. (113). With the spot continuously moving, the effect of the ions is spread more or less uniformly over the screen so that the decrease in luminous efficiency is not great. However, where the beam is electrostatically focused and magnetically deflected, the ions and electrons reach the deflection system moving together on or near the axis of the tube. The angle of deflection, as shown by Eq. (114), is inversely proportional to the square root of the mass, so that, for instance, an oxygen ion O_2^- is deflected only by an amount $d/\sqrt{32} \times 1,850$, which would be about 0.01 in. when the spot due to the electron beam is displaced 3 in. The result is that in time a dead spot, or *ion burn*, occurs in the center of the screen.

Ion burns are prevented in tubes employing magnetic deflection when an ion trap gun is used and are inhibited by the use of an aluminum screen backing.

In an ion trap gun the beam is first formed, then bent by means of a constant magnetic field. The electrons are then allowed to pass on toward the screen, while the ions, which were substantially undeflected, are captured in the electrodes or on the neck of the bulb. This requires that the tube be provided with a bend in the gun or neck of the bulb.

It is evident, however, that if a magnetic field and an electric field are applied at right angles to each other at substantially the same location on the axis, the ions may be separated from the beam when the fields are adjusted to produce equal but opposite deflections for electrons by Eqs. (113) and (115). Although such an arrangement is not readily achieved in practice, guns have been made in which the electron path does not move far off the axis of the tube and the ions are substantially all removed from the beam.

52. Aluminum-backed Screens. The aluminum screen backing consists essentially of a very thin film of aluminum laid flat over the screen. Its principal purpose is to reflect the light emitted by the phosphor so that it is passed back through the face of the tube where it may be seen, thus improving the over-all efficiency. While the loss in energy of an electron in passing through the aluminum backing may be of the order of a few thousand volts, the over-all efficiency of the aluminum-backed screen is

greater than the unbacked screen at beam potentials greater than about 5,000 volts. The aluminum backing has an inhibiting effect on ion burning.

53. Phosphors. The coating of the screen in a cathode-ray tube is called a *phosphor* and the process by which a part of the energy of the bombarding electrons is converted to light is called *fluorescence*. When the beam is cut off, the light output usually decays rapidly at first, then more slowly because of *phosphorescence*. These phosphors ordinarily consist of complex sulfides or silicates *activated* with traces of manganese, silver, or other (generally metallic) elements. A short description of some of the more common standardized phosphors follows.

P1. Zinc orthosilicate (Zn_2SiO_4) manganese-activated. This phosphor is commonly known by its crystallographic name, willemite. It produces green light peaked about 5250 Å and is widely used in tubes intended primarily for oscillographic work. When excitation or bombardment is discontinued, the brightness of the spot decreases to 10 per cent of that obtained during excitation in about 0.017 sec, which is to say that the phosphor has *medium persistence*.

P4. Commonly, silver-activated zinc sulfide mechanically mixed with manganese-activated zinc beryllium silicate. Silver-activated zinc cadmium sulfide is sometimes used with, or in place of, the zinc beryllium silicate, the silver-activated zinc sulfide being a constituent in any case. To the average human eye the color is white, so that this phosphor is widely used in television tubes. Its luminous efficiency is comparable with that of the P1 phosphor; it also has medium persistence.

P5. Commonly, calcium tungstate, sometimes containing lead. The color is a deep blue, so that it is particularly well adapted to oscillographic work where the curves are to be photographed. Its luminous efficiency is about half that of the P1 or P4 phosphors, and its persistence is much shorter.

P7. The P7 screen is produced by laying a copper-activated zinc cadmium sulfide phosphor on the faceplate of the tube, then depositing silver-activated zinc sulfide over the first layer. In operation, the inner phosphor is excited by the electron beam with the emission of blue light which, in turn, excites the copper-activated layer next to the glass with the emission of yellow light.

The outer layer of phosphor has the peculiar property of being rather slow to emit after excitation is begun by the blue light of the inner layer. It is this property which makes the P7 screen of particular importance in radar work. When viewed through an amber screen which filters out the blue light, only the pips or other repeated signals are seen, the noise and other nonrepetitive signals being masked out.

P11. Commonly, silver-activated zinc sulfide to which a small amount of nickel has been added. This is a short-persistence screen of good efficiency giving a blue color well adapted to oscillographic work, especially where photographic records of the trace are required.

P12. Commonly, manganese-activated zinc magnesium fluoride. The allowable current-density limit is lower than for most of the other screens, but the exponential decay characteristic of the yellow light gives it special significance in the observation of repetitive signals in instrument and radar applications.

Phosphors may be applied to the face plate of the tube by spraying a suspension of the phosphor in a thin lacquer. The preferred procedure is to suspend the phosphor in distilled water and drop a measured quantity of the suspension into the bulb by means of a pipette, then siphon off the liquid after the solid has had sufficient time to settle out, the time required being of the order of 1 hr. The remaining liquid is then evaporated at a convenient temperature not over 100°C, followed by baking at a higher temperature. During exhaust, the bulb will again be baked to remove occluded gases as for any kind of tube. For a cathode-ray tube this baking is commonly done at a temperature range of 360 to 400°C for about 15 min. The amount of phosphor used depends to a certain extent upon the variety; for example, in using the P1 phosphor the average weight of dry powder used is about 5 mg per sq in.; for P4, the average weight is 26 mg per sq in.

CHAPTER 14

ANTENNAS

BY EDMUND A. LAPORT¹

The transmission and reception of electromagnetic waves used for radio communication are accomplished by radiators and collectors exposed in space and known as *antennas*. An antenna is a device composed of a system of one or more linear conductors, usually of large electrical dimensions, from a fraction to several wavelengths, which is used to couple a h-f a-c generator or receiver to space. Between the transmitting and the receiving antennas there is a combination of earth, water, air, and ionospheres which constitute the mediums in which electromagnetic waves are propagated. The action of the waves in traversing these mediums is very complex, being dependent upon many known and other unknown factors. Prominent among the known factors are the transmitting frequency, the radiation characteristics of the transmitting antenna, the orientation of the path of transmission in the earth's magnetic field, the time of day and the conditions of daylight and darkness along the path, the season of the year, the solar activity, the electrical characteristics of soil or water in the immediate vicinity of the antenna as well as along the path of the surface waves, the immediate conditions of ionization of the atmosphere at various levels, the distance between transmitter and receiver, and the characteristics of the receiving antenna.

The whole field of antenna technology has grown from the striving to launch and collect radiant energy in the most efficient ways dependent upon the empirical facts of wave propagation. If the earth were a smooth perfectly conducting sphere in a vacuum, antenna techniques, except for size, would not vary much over the entire useful range of radio frequencies. In reality the lowest frequencies are best utilized by ground-wave transmission. These gradually blend into the frequencies where both ground waves and sky waves have separate usefulness. Gradually these blend into the frequencies for which ground waves are rendered useless and only sky waves can be usefully employed. Eventually frequencies are reached where the ionospheres no longer reflect waves and where range is limited to the diffraction region beyond the horizon. At still higher frequencies, this diffraction zone is reduced to virtual uselessness, and transmission is confined to purely optical circuits, but complicated sporadically by meteorological effects which produce mirage. Meteorology plays a greater part in wave propagation as the frequencies enter the s-h-f region. These different portions of the spectrum each require different basic antenna techniques, and in each portion a large variety of antenna structures has been evolved with different degrees of general or special utility.

It must be emphasized repeatedly that antenna applications are dependent upon the nature of wave propagation for any frequency and transmission circuit, and for that reason the antenna engineer must also be expert in propagation physics.

It is impossible to place on paper all the various considerations that determine the design details of an antenna. For different applications different weights must be placed on each factor in the complex combinations of electrical, mechanical, and economical aspects of the design. Even though the same fundamental principles apply generally, a carefully engineered antenna system is an empirical entity, calling for a certain seasoned judgment gained only from experience or close study of the theory and techniques developed for various purposes throughout the years. But

¹ Chief Engineer, RCA International Division, Radio Corporation of America, New York.

there is no intrinsic mystery to the subject, and the fine points merely provide those extra few decibels of superiority; and whether this superiority is worth what it sometimes costs, the designer must decide for the particular case. As in most fields, over-design may demonstrate poor judgment as much as underdesign.

In general, it is possible to excite any antenna at any frequency simply by applying enough potential to drive a current against the impedance it presents at that frequency, whatever it may be; or to match this impedance into that of the generator so that all the power is delivered up to the point of termination. Beyond this point, the power must either be consumed in internal losses in the system or be radiated. But to radiate the power in a particular manner into space for effective utilization, the antenna must be designed to have current distributions in the radiator which shape the fields in a desired, as distinguished from a random, manner.

1. **Antenna Terminology.** The following terms are used in this work:

Meter-amperes. In general, this means $\int i dl$, where i is the rms current in an elementary length of the antenna, dl . The integration is performed over the entire vertical length of the exposed (radiating) parts of the radiator. Viewed geometrically, this is the area of a plot of rms antenna current in amperes against distance along the antenna measured in meters. The directions of the currents must be considered.

Doublet. A differential of antenna length, short enough to be considered to have uniform current throughout its length.

Dipole. (1) *Electric Dipole.* A radiating element producing a field in which the electric vectors lie in planes passing through the axis of a linear charge distribution, and characterized by a linear radiator. A half-wave ($\lambda/2$) linear radiator is usually called a half-wave dipole.

(2) *Magnetic Dipole.* A radiating element producing a field in which the magnetic vectors lie in planes passing through an axis normal to a closed distribution of currents, and characterized by a loop antenna.

Self-impedance. The impedance of a single radiating element in the absence of any influences from other radiators, as measured at a current antinode or other feed point. It is the ratio of the impressed voltage and the current.

Mutual Impedance. The circuitual equivalent of radiation coupling. Mathematically expressed, it is the negative ratio of the induced potential at the base (or the current antinode) of a second radiator to the base current (or antinode current) of the first radiator.

Harmonic. Any natural frequency of oscillation of a system expressed as a number which is the multiple of the fundamental frequency.

Antenna Tuning. The act of resonating an antenna system to some frequency other than a natural frequency by means of reactive devices.

Antenna Loading. Lumped reactances connected in the antenna system for the purpose of antenna tuning.

Distributed Loading. Units of reactance added at small electrical intervals along a conductor for the purpose of smoothly modifying the natural distributed constants of the system. Pupinization.

Node, or Nodal Point. In a standing-wave system, the points of either zero or minimum potential or current.

Antinode. In a standing-wave system, the points of maximum potential or current.

Vertical Polarization. A wave orientation such that all the lines of electric force lie in planes perpendicular to the ground plane, in the direction of propagation.

Horizontal Polarization. A wave orientation such that the lines of electric force are parallel to the ground plane.

Circular Polarization. A wave in which the direction of displacement at a point rotates about a point in a plane with constant magnitude such as to describe a circle. Commonly known as a rotating field. When the displacement rotates asymmetrically and its magnitude varies as the radius of an ellipse, the wave is said to be *elliptically polarized*.

Reflector. Conductor or conductors so disposed with respect to a radiator as to react upon the latter in a manner which transforms the radiation pattern by suppressing radiation in its direction while reinforcing it in the opposite direction.

Antenna Array. A multiplicity of radiating elements disposed in any manner whatsoever for the purpose of molding the space characteristic in some desired fashion.

Radiation Pattern. A means for describing the over-all radiation characteristics of an antenna system. Usually refers to a geometrical solid in spherical coordinates with

distance from the origin proportional to the radiation intensity in any direction. Radius vectors may be proportional to field intensity or to power. Power flow by radiation in any direction is proportional to the square of field intensity.

Fundamental Frequency. The frequency at which the impedance of an antenna at a current antinode is minimum. The lowest frequency of self-oscillation of an antenna.

Fundamental Wavelength. The length of the space wave emitted by an antenna oscillating at its fundamental frequency.

Mode of Operation. The ratio of the operating wavelength to the fundamental wavelength; also, the ratio of the fundamental frequency to the operating frequency.

Electrical Length. The length of a standing wave in any linear system expressed in degrees or radians. The electrical length of a wire is its actual length in terms of wavelengths and fractions thereof multiplied by 360 deg.

Effective Height. The height h obtained from the following equation:

$$h = \frac{ed}{1.25fI} = \frac{\text{meter-amperes}}{I_0}$$

where h = effective height, m

e = measured field intensity, microvolts per m

d = distance from antenna to point where e is measured, km

f = frequency, kilocycles

I = antenna current at point where antenna is energized

NOTE: d must be small enough so that the effect of attenuation is absent, and great enough to be beyond the limits of the induction field.

Antenna Resistance. The total dissipative component of the antenna impedance measured at the point where power is introduced.

Radiation Resistance. The ratio of the total power radiated by an antenna and the square of the current at some reference point in the system, usually the point where power is introduced, or a current antinode.

Standing- (or Stationary-)wave Antenna. An antenna or antenna system in which the current distributions are produced by standing waves of charges on the conductors.

Traveling- (or Progressive-)wave Antenna. Antenna in which the current distributions are produced by waves of charges propagated in one direction only in the conductors.

Band Width of an Antenna System. The total number of cycles included within the points of equal response of the frequency selectivity curve as seen from the feed point of the system. In practice this is often taken as the frequencies at which the impedances depart from resistance by a certain amount in terms of power factor; or, in transmission-line coupled systems, the departure of the feeder standing-wave ratio (SWR) from unity by some specified amount. There has been no general standardization for the limits at which band width is reckoned.

Beam Width. The total angle (in degrees or radians) included within points of equal power or field intensity of the main lobe of a radiation pattern. There has been no general standardization of the specification for beam width, but one method used is the half-power beam width (field intensity equals 0.707 of the maximum), and the quarter-power beam width (field intensity equals 0.5 of the maximum).

Main (Principal, Major) Lobe. The largest lobe of a radiation pattern and the one utilized for the intended application for the antenna. Some radiation patterns have more than one main lobe.

Secondary (Parasitic, Minor) Lobes. All the smaller lobes of a radiation pattern; usually having no utility.

2. Radiation Control. Almost all modern radio applications require some specified form of radiation control for power gain, antifading, reduction of interference, radio guidance, reduction of multipath distortion, or reduction of noise. Antenna technology plays an important role in the general struggle to improve utilization of the radio spectrum.

Radiation is controlled by setting up current distributions which produce wave interference patterns in space having specified or desirable forms best adapted to the wave propagation circumstances. Generally, precise control in three dimensions is necessary. The higher the frequency, the easier it becomes to achieve high degrees of radiation control, limited only by frequencies so high that structures of conductors become too small to be practical and one must resort to horns. At the l-f end of the spectrum, the limitations are due primarily to great size (and cost).

The principles of wave interference have been highly developed in electromagnetic theory, concerning which there is a vast literature with many excellent recent contributions.¹ The principal advances in bringing this theory into intimate engineering use have been the application of practical units through the mks system, and the concept of the characteristic impedance of free space. The abstractions of the Gaussian system of units have been somewhat a handicap to general practical use.

3. Current Distributions in Linear Conductors. The treatment of an antenna as an open-ended uniform transmission line with uniformly distributed attenuation is not in full accord with modern physical theories. However, this viewpoint is the simplest approach to a very complicated subject and is sufficiently in accord with practical results to justify its presentation in this form in a treatise of this type.

The various conductors of a radiating system carry currents of varying amplitudes, phases, and configurations in three dimensions. The fields set up by all the infinitesimal current elements of a system, together with the reflections which occur from the earth and reflecting areas, produce a characteristic radiation pattern. The problem in design is to set up current distributions which produce a desired radiation pattern. For this purpose, two basic types of distribution may be mentioned: those derived from standing waves on the conductors, and those derived from traveling waves. The first may be coarsely said to be (for a conductor length less than $\lambda/2$) a current distribution of constant phase and variable amplitudes. The second is essentially a constant amplitude and variable phase distribution. It aids understanding of design principles to recognize these differences, even though practical systems have both characteristics in varying proportions.

In addition to their effect on the special distribution of radiant energy, current distributions determine the circuit impedance of an antenna, and the engineer must have regard for this fact for a number of reasons. Many forms of radiation patterns are realizable only by designing for predetermined impedance relationships at various points in the antenna and feeder system. As antenna techniques are developed, it becomes necessary at times to depart from simple natural current distributions and set up special distributions in linear conductors, sheet conductors, and loops. Antennas are really a complex of radiating portions and feeder portions, which have to be considered together. Only in the most elementary systems can the radiators and the feeders be treated separately.

4. Stationary- (Standing-)wave Distributions. On a thin uniform linear conductor of arbitrary length with its far end open, excited at one end by an a-c generator, the current distribution is a result of interference between a wave of charges traveling outward from the generator and another wave, due to reflection from the open end, traveling backward to the generator. The resulting distribution is a standing wave of currents following the expression,

$$i_x = f(x) \doteq I[e^{(\alpha+j\beta)x} - e^{-(\alpha+j\beta)x}] = I \sinh(\alpha + j\beta)x$$

As $\alpha \rightarrow 0$,

$$i_x \rightarrow I \sin \beta x$$

where i_x = current amplitude and phase at a distance X from open end of a conductor of length l , including ΔX , an apparent elongation due to end effects

α = attenuation constant, nepers per unit length (1 neper is 1 hyperbolic radian = 8.686 decibels)

β = phase constant per unit length, cir radians $2\pi/\lambda$, λ being the wavelength in conductor

$$I = I_0 e^{-\alpha x}$$

where I_0 is current entering conductor from generator if conductor were of infinite length, and having a characteristic impedance.² For vertical grounded radiator

¹ See references 1, 2, and 3 at the end of this chapter.

² See reference 15.

$$Z_0 \doteq 138 \log_{10} \frac{2h}{d}$$

The physical significance of the above is as follows: A current $I_0 = E_0/Z_0$ enters the conductor and describes a logarithmic spiral of decrement α and negative phase rotation β (determined by the velocity of propagation of charges) as it passes along the wire, gets inverted at the end, and returns to the generator. The vector sum of the going and returning currents at any point X is i_x . A comparison between theoretical sinusoidal and measured distributions on thin wires is shown in Fig. 1.

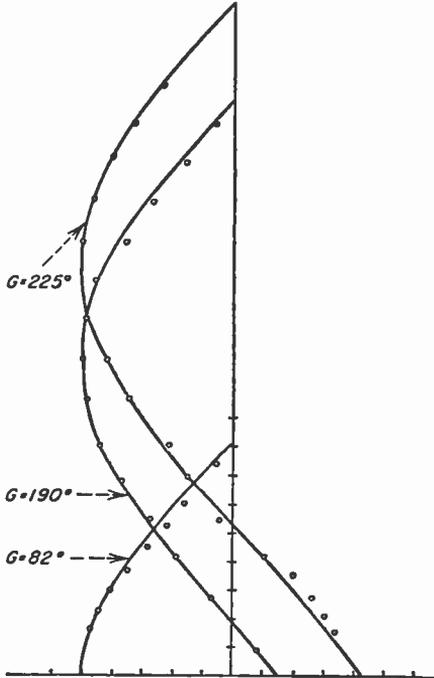


FIG. 1. Current distribution on vertical antennas of three lengths and small cross section. Circles are measured values; solid lines, true sinusoidal distribution.

phase distribution, especially in the nodal regions, and e_x has a phase relationship with respect to i_x which is always less than 90 deg and comes into phase with i_x every quarter wavelength from the end.

Owing to the end effect ΔX , the distance to the first potential minimum is always slightly less than $\lambda/4$ from the open end. The amount varies from 2 per cent for thin wires to 10 per cent for large-diameter conductors and for the capacitive effect of insulators.

The impedance at any point in the system, looking toward the open end, is the vector ratio $Z_x = e_x/i_x$. $Z_x = R \pm jx$ is called the self-impedance as seen from that point (usually of interest only at a feed point).

Experimentally determined curves of resistance and reactance of cylindrical antennas normal to a perfectly conducting ground are shown in Figs. 2a and b.¹ For isolated balanced antennas the values shown must be doubled. These are shown

¹ See reference 17.

In the theoretical case of $\alpha = 0$, the amplitude of i_x will vary sinusoidally with distance X , and remain constant in phase except for successive reversals at points which are multiples of $\lambda/2$ from the end. In reality, radiation and heat losses give finite values to α which are small normally. This causes i_x to depart slightly from a sinusoidal distribution, especially in the vicinity of the nodes, and causes the phase to change slowly with distance except in the vicinity of the nodes where it rapidly changes its phase to substantial reversal. In all but the most exacting cases, the current distribution on thin uniform linear conductors can be assumed sinusoidal for purposes of calculating radiation patterns.

In the same manner, the potential distribution along the conductor can be written

$$e_x = f(x) \doteq E[e^{(\alpha + j\beta)x} + e^{-(\alpha - j\beta)x}] \\ \doteq E \cosh(\alpha + j\beta)x$$

In the ideal case of $\alpha = 0$, the potential distribution is cosinusoidal in X , reverses its phase at values of X which are odd multiples of $\lambda/4$, and e_x is always in quadrature to i_x at any point. For small finite values of α , e_x departs slightly from this ideal amplitude and

in terms of the parameter D , which is the diameter of the antenna in electrical degrees, and include the capacitance of the closed lower end of the cylinder which in all cases is 1 deg from ground. These curves have fundamental reference value.

Empirical Distributions. By modifying the velocity of propagation and the characteristic impedance of the conductor, uniformly or nonuniformly, or by use of lumped reactances inserted in the system, new expressions for the current and potential distributions are required. Empirical distributions can be produced in this way.

5. Traveling- (Progressive-)wave Distributions. If instead of being open the end of this conductor is terminated so as to absorb all energy arriving at the end

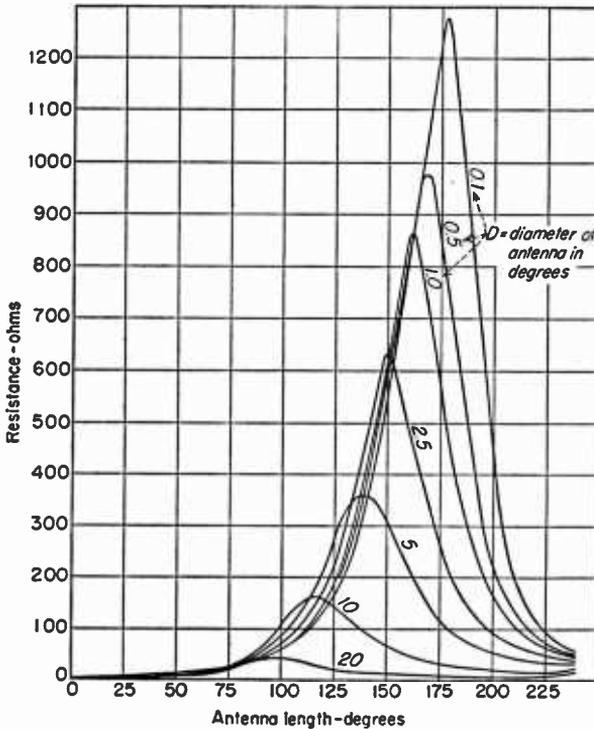


FIG. 2a. Measured resistance of cylindrical antenna normal to perfect ground.

without reflecting any of it, then a traveling-wave system results. Under perfect realization of this condition $Z_x = Z_0$ is constant for any value of X .

In the standing-wave condition, impedance variations with length vs. frequency exhibit properties analogous to resonance ($X = \text{odd multiples of } \lambda/4$) or antiresonance ($X = \text{even multiples of } \lambda/4$). These resonance effects are selective to frequency in varying degrees and limit the band widths that can be used. In the traveling-wave condition, selectivity is impossible unless the termination is selective.

6. Special Distributions. Arrays of conductors having predetermined current distributions, amplitudes, and relative phases of excitation, produce special distributions of current in three dimensions for radiation control. By so doing, each conductor is immersed in the fields of all the others, producing energy interchanges through radiation couplings. They produce circuitual effects called mutual impedances. Depending on spacing, relative current amplitudes and phases, and orienta-

tions, the mutual impedance between two radiators at a given frequency is in general complex and can be vectors lying in all four quadrants. R and X can both be either positive or negative.¹

7. Distributed Potentials Induced by Space-wave Fields. When one conductor is in the field of another, the energy exchange is by means of potentials induced in all parts of the wire. These potentials would have to be simulated by a system of dis-

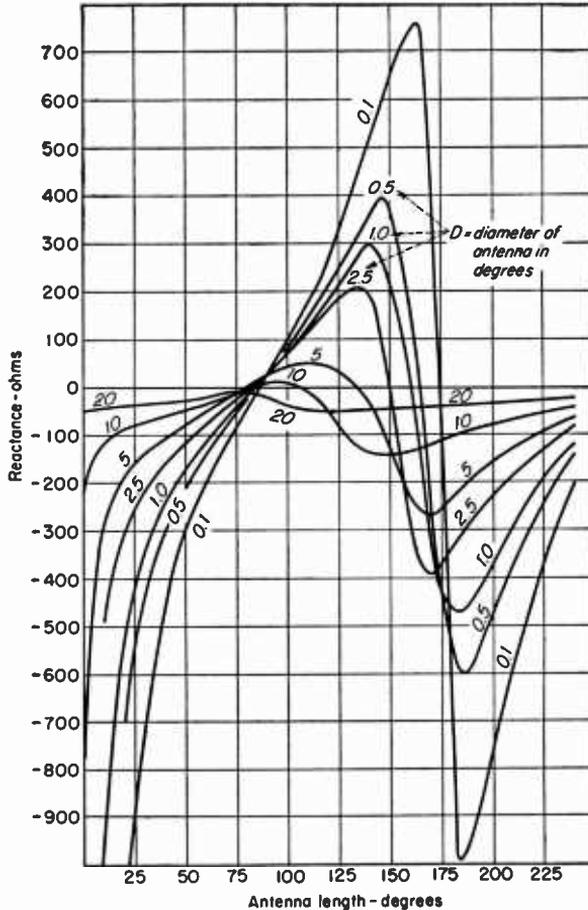


Fig. 2b. Measured reactance of cylindrical antenna normal to perfect ground.

tributed generators in contrast with a single generator used to excite a single radiator. Distributed generators produce a different current and potential distribution along a wire than does a single generator such as a transmitter. Therefore, mutual impedance effects modify the shape of the current distributions from those prevailing in the absence of mutual impedance.

In the case of receiving antennas, which are excited entirely by the field of the passing electromagnetic wave, the current and potential distributions are not like that for transmitting except for a $\lambda/4$ grounded antenna and a $\lambda/2$ antenna above ground.

¹ See references 15, 19, 36, 42, 44, 45, 53, 54, 62, 69, 102, 107.

LOW-FREQUENCY TRANSMITTING ANTENNAS (BELOW 300 KC)

With the advent of high frequencies for point-to-point communications during the last two decades, the low frequencies were relegated to the past in the popular mind. There has been a vigorous revival of the use of low frequencies for transmission in the auroral zones where the high frequencies are too unreliable for certain important

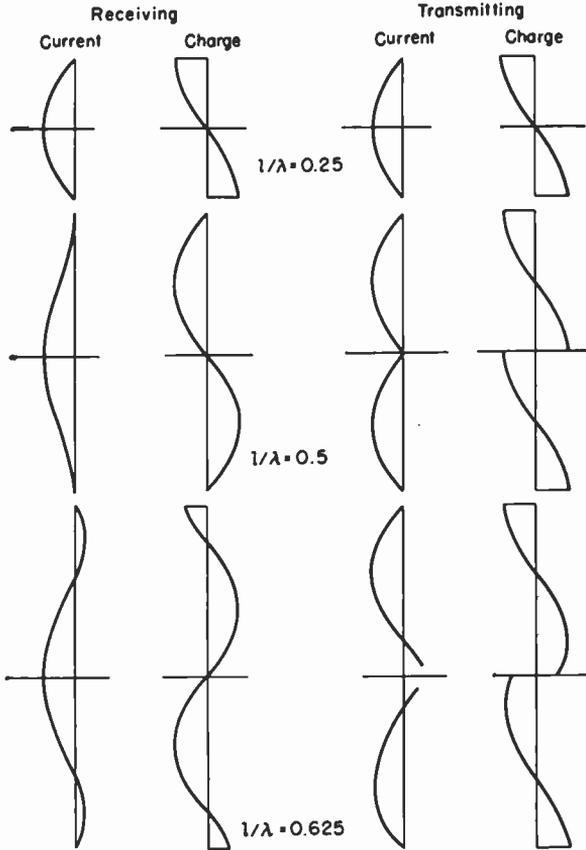


FIG. 3. Current and charge distribution in thin uniform wires for conditions of free oscillation.

services such as aviation. Fortunately the auroral zones are among the lowest noise zones of the world so that, aside from the factor of propagation, the reliability and the effectiveness of low frequencies are much greater than for the high noise zones of the lower latitudes.

V-l-f and l-f antennas are generally characterized by short electrical length with consequent low radiation resistance and high input reactance. Design techniques feature methods for maximizing radiation resistance and minimizing ground resistance, insulation loss, and reactance. Power input is limited by corona or flashover potentials. The potentials on the system are uniform within a few per cent. Useful radiation is that which is vertically polarized, and service is rendered by ground wave with some D-layer influence. Ground system design must take into account the

need to minimize current densities at collecting points and the very deep penetration of ground currents into soils of highest conductivities and even into sea water. This is the oldest branch of antenna technology, and yet there is room for additional research.

8. Fundamental Wavelength. One of the first antenna characteristics to be selected or determined is the fundamental wavelength (λ_0). From this, electrical length (G) is derived for any operating wavelength (λ) and applied to various other formulas for radiation resistance, reactance, and potential. At its fundamental wavelength an antenna is said to be 90 deg ($\lambda/4$) long.

For an antenna consisting of a supported vertical wire or a vertical tower of uniform cross section the following formula applies within approximately 4 per cent:

$$\lambda_0 = 1.33h$$

where λ_0 is in meters and h is in feet.

For symmetrical T antennas the fundamental wavelength can be predicted only approximately from simple formulas, and for large-area flattop systems it usually must be estimated, or measured from scale models. Many empirical factors are the cause of this uncertainty, including the ratio of horizontal to vertical lengths, the capacitance of insulators, the proximity of grounded or insulated towers, the number of conductors used and their cross-sectional configuration, the reflections from the T junction, and perhaps other factors. For T antennas, one can use the formula

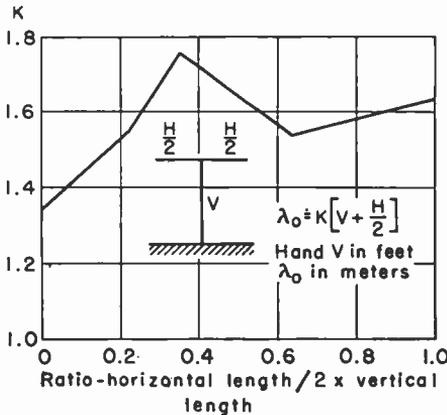


FIG. 4. Empirical values of fundamental wavelength (λ_0) for T antennas (from measurements of 24 antennas).

$$\lambda_0 \doteq k (\text{vertical height} + \text{one-half the horizontal length})$$

where the antenna dimensions are in feet, the wavelength in meters, and the coefficient k is taken from Fig. 4. This figure was derived from 24 actual antennas, which were carefully measured. The antennas were of different sizes and with several different tower arrangements. This unusual curve accounts within 5 per cent for all but four of the antennas. The corners in the curve are based on reliable data and so were not rounded out. The value at the right-hand extremity of the curve is based on three identical 1-f antennas constructed in three different locations and measured by separate qualified individuals with exact concordance.

9. Radiation Resistance. The radiation resistance is increased by increasing the vertical height and by producing, as nearly as can be realized, a uniform current distribution on this portion of the antenna. This is usually done by some capacitance loading at the upper end in the form of an inverted L, a T, or large flattop. The radiation resistance can be calculated from the following relationship:

$$R_r = \frac{30}{\sin^2(A+B)} \left[\frac{4A^2}{3} \sin^2 B + \frac{2A^2}{3} \sin 2B + \left(\frac{17}{30} \cos 2B - \frac{7}{30} \right) A^4 \right]$$

where A = electrical height of vertical section

B = equivalent electrical length of flattop

R_r = radiation resistance, ohms

When $B = 0$, $R = 10A^2$ radians (which holds for a vertical antenna).

The above relationships are valid for $A \geq \lambda/12$ ($G < 30^\circ$).

The electrical length of an inverted L or a T antenna can be taken, for all practical purposes, as the sum of the vertical and horizontal parts in electrical degrees (or

radians) provided that the horizontal distances from the junction with the vertical are electrically small. For a capacitance area of empirical form, calculation may be difficult. In such cases, it is desirable to use a scale model, the scale being in size directly proportional to wavelength and erected over a metallic plane of area very large with respect to that of the flattop. The supports should also be modeled and the locations of antenna and guy insulators simulated. For reactance measurements the cross section of the antenna conductors should also be to scale.

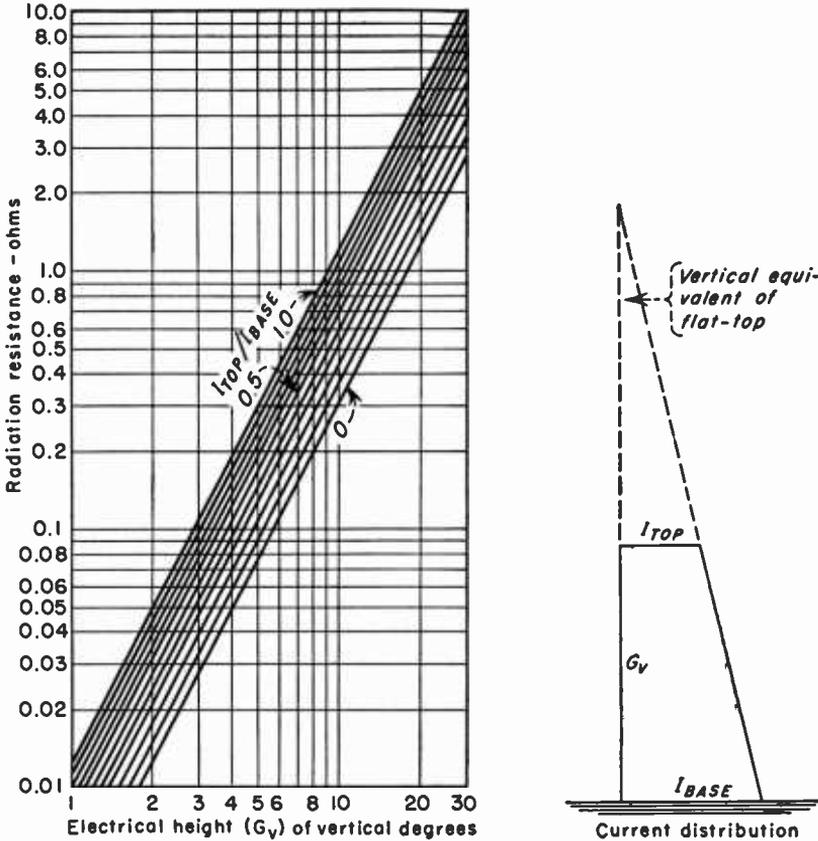


FIG. 5. (Left) Radiation resistance for electrically short antennas vs. height of vertical portion (degrees) and (right) ratio of currents at top and base of vertical portion.

It is a great convenience to determine l-f antenna performance in terms of meter-amperes in the vertical portion. In most l-f antennas, the current distribution on the downlead is linear, and the ratio of current near the flattop to that near the base can be accurately calculated from the electrical dimensions, permitting in turn an accurate calculation of the meter-amperes provided the current can be observed at the lower end of the downlead, above all tuning apparatus. Field intensity is directly proportional to meter-amperes at any one frequency. Since a perfectly efficient $\lambda/12$ vertical antenna with 1,000 watts radiated gives a field of 187 mv per m at 1 mile, an excellent reference for determining radiated power is available. (This reference is used because longer vertical antennas begin to provide some gain by vertical direc-

tivity.) Then if the total antenna resistance can be measured and the total power input determined, the radiation efficiency can be computed from field strengths (see Fig. 6).

The electrical length of an antenna is found from the relationship

$$G = \frac{\pi\lambda_0}{2\lambda} \quad \text{radians} \quad \text{or} \quad G = 90 \frac{\lambda_0}{\lambda} \quad \text{deg}$$

The electrical length of the vertical and that for the flattop can then be segregated for use in the above equation for radiation resistance.

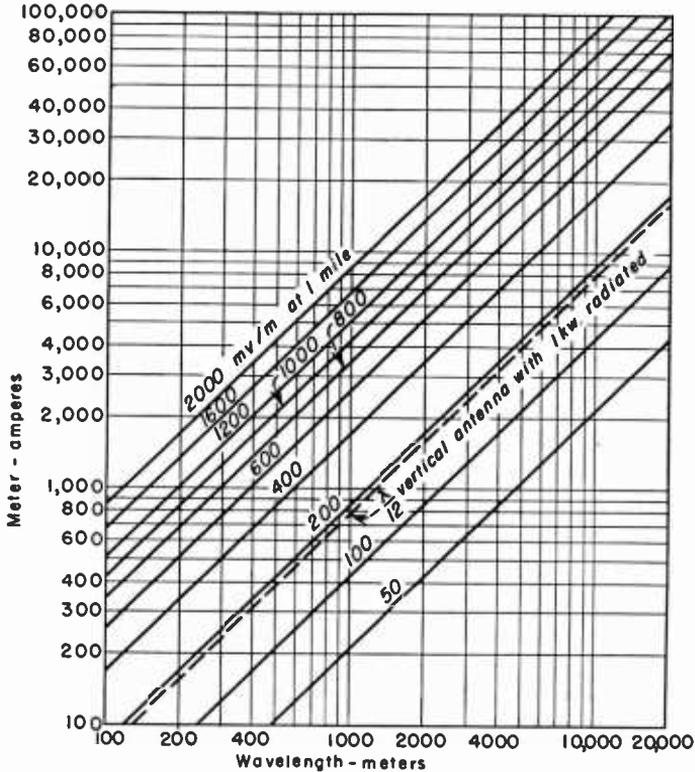


FIG. 6. Meter-amperes required to produce given field at given wavelength.

The maximum field intensity possible with a given power input is the objective of all antenna designing. With l-f systems the cost per decibel of improvement at times may be very great, especially at the lowest frequencies.

The measurement of antenna resistance of antennas of very short electrical length is a difficult operation and requires special power oscillators to permit the use of thermocouple ammeters having internal resistances small with respect to that of the circuit being measured. The resistance-variation method is also difficult unless special heavy-duty standard-value resistors are provided. The reactance-variation method is difficult because of the range of variability required in a heavy-duty tuning device. The r-f impedance bridge is an excellent instrument for such measurements.

The most convenient method of measuring antenna performance is by field intensity

measurements at standard distance. Since large antennas which are multiple-tuned are not point-source radiators, substantial distances are required for observing performance of the v-l-f systems. For antennas with single tuning, distances as close as 1 wavelength may be used for measuring the radiation field. For wavelengths shorter than 1,600 m, 1 mile is a desirable standard distance, measured from the download.

10. Scale-model Measurements.¹ Valuable data on a l-f antenna can be obtained quickly and economically by h-f scale-model measurements. Two very important values, reactance and fundamental wavelength, can be determined accurately using laboratory instruments such as a Q meter or other r-f impedance-measuring device.

Resistances of ground and conductors and insulators cannot be simulated. It is, therefore, desirable to build the scale model over a large continuous metallic plate acting as a perfect ground. The linear and cross-sectional dimensions are to the same scale. Towers and guys can be simulated precisely, and the effects of placement of insulators in supports and guys can be measured.

The fundamental wavelength can be measured by connecting the download to the ground plate and shunt-exciting the system with very loose coupling from a buzzer. The latter must be shielded. The emitted wavelength is measured by reception using a calibrated receiver. Also any of the familiar techniques of measurement of resonance in the antenna with a coupled oscillator may be used if preferred. The buzzer method is very simple and satisfactory.

Reactance can usually be measured over the range of values of interest using a Q meter. For conditions of high Q values, body effect near the antenna produces errors which can be eliminated by manipulating the Q meter from a distance. Since capacitive reactances are being measured, appropriate methods must be employed.

Scales as extreme as $\frac{1}{100}$ of natural size have proved successful with careful measuring techniques. Scales of the order of $\frac{1}{5}$ are more easily measured, but the model may be very large. The ease with which a number of empirical constructional details can be measured and compared permits the final design to be quite thoroughly evolved in the model stage. The potentials, insulator arrangement, conductor arrangements, rigging details, support locations, tower heights, etc., can be developed as far as desired.

The principle of similitude upon which model tests are based can also apply in the opposite direction by using, for instance, v-h-f models for microwave antennas, where the enlarged antenna may be more easily measured and developed than at natural size. In all scale-model work, the frequency used is proportional to the scale of the model. At $\frac{1}{50}$ scale, 50 kc is simulated by 2.5 Mc.

11. Reactance. The reactance of the antenna at the operating wavelength λ is an important value to determine so as to know the required size of the antenna load coil (l-f antennas always have a negative reactance) and the antenna potentials encountered. The reactance must be minimized to reduce loading inductance (and load coil resistance) and to reduce insulation loss even when potentials are far below that which produces corona. This also reduces the selectivity of the antenna system, permitting higher keying speed without excessive telegraph distortion, or greater frequency range for telephony. Reactance is reduced by increasing the top-load capacitance, increasing the cross section of the antenna with cage or flat multiple conductors, or both. Considering the antenna as a transmission line with characteristic impedance $Z_0 = \sqrt{L/C}$,

$$X = -jZ_0 \cot G = -j138 \left(\log_{10} \frac{l}{\rho} \right) \cot G$$

where X = ohms

l = length of straight antenna

ρ = radius of antenna in same units as l

G = electrical length

These equations are satisfactory for $G \gtrsim 30$ deg.

¹ See reference 15.

It is seen that $\cot G$ is important but is determined by the electrical size of the antenna and has a large influence on its cost. For a given electrical length and size of structure, $\cot G$ is constant. Further reduction of reactance must, therefore, be made by reduction of Z_a , which is done by decreasing inductance and increasing capacitance per unit length. There are mechanical limitations to cage or other forms of multiwire conductors because of wind and sleet, the increased sags and tower loadings, and the multiplied problems of vibrational stress and fatigue. The curves in Fig. 7 show the increase in capacitance per unit length using cages in terms of percentage increase over a single wire for a particular case, but represent a general case qualitatively. One can choose a point where mechanical objections outweigh electrical advantages.

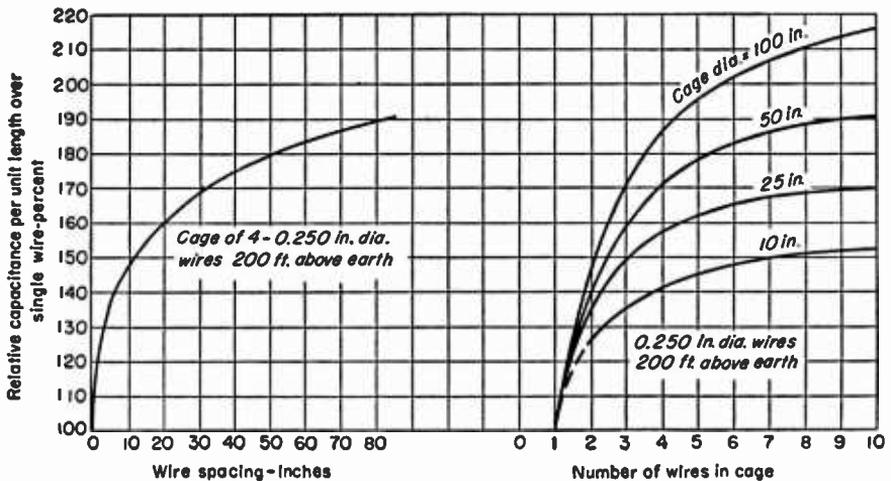


FIG. 7. Increase in capacitance of cage vs. single wire.

12. Antenna Potential. The potential E_a at the feed point of the antenna (near ground) for a power input W into the antenna impedance $Z_a = R_a - jX_a$ is

$$E_a = I_a Z_a$$

The antenna current $I_a = \sqrt{W/R_a}$.

Most l-f antennas (or any electrically short antennas) have a reactance large with respect to the resistance, so that usually

$$E = I_a X_a$$

The antenna potential is kept to a minimum for practical reasons of simplifying the insulation and avoiding corona or flashover. Therefore the reactance of the antenna is made to be as low as practicable for a given total antenna resistance and power input.

There is a rise in potential along the antenna above the feed point, the amount varying with the electrical length of the system and upon the configuration of the conductors. For a uniform vertical antenna, the maximum potential becomes

$$E_{\max} \doteq \frac{E_a}{\cos G}$$

In the case of an antenna with a very high capacitance flattop,

$$E_{\max} = \frac{E_0}{\cos A}$$

Where A is the electrical height of the vertical section.

Prevention of corona, pluming, and flashover is a matter of both conductor and insulator design. Potential gradients must be kept below critical values for the ionization of air at the altitude of the site, under all meteorological conditions. Moisture films on insulators and drip water on conductors often increase losses or cause corona or plume formation, especially in still air. Adequate allowance has to be made for insulation in any case, but when antenna potentials surpass 100 kv, very special design problems are involved even at sea level; if altitudes of 5,000 to 10,000 ft are encountered at the site, the power may be severely limited, or the fundamental design of the system is affected. In the latter case it is simpler to employ self-supporting towers as radiators to attain low gradients for relatively high potentials by virtue of the large cross section of the tower. This centers the flashover problem in the base insulation where it is not too difficult to take the required precautions for satisfactory design.

13. Antenna Resistance. The electrical design of a l-f antenna must maximize radiation resistance and minimize conductor, insulation, and ground components of resistance. In the antenna circuit is also the tuning inductance, the resistance of which must also be minimized. In a well-designed system, ground and tuning coil resistances are the principal sources of serious power loss, and special attention must be directed to these matters to obtain maximum radiation efficiency.

14. Multiple Tuning.¹ V-l-f antennas often have radiation resistances which are only a fraction of an ohm. To obtain a radiation efficiency even as high as 6 per cent it has been necessary to use several downleads from the flattop and tune each downlead separately. The antenna is fed from one point only. This technique of multiple tuning was an outstanding development in the improvement of v-l-f antenna efficiency. The use of a separate ground system under each downlead reduces ground-current densities and, therefore, total ground resistance, because the total antenna current is divided among the several ground systems. The radiation resistance as seen from one downlead increases as n^2 where n is the number of multiple-tuning downleads.

15. Ground Systems. For l-f working, the reduction of ground resistance starts with the selection of the site, which should be an area of highest available conductivity to a considerable depth and for a distance of more than $\lambda/2$ from the antenna. The site need not be topographically flat although an undulating surface complicates construction. Hills or mountains surrounding should have an altitude which is a small part of a wavelength to avoid shadows and high attenuation for ground-wave propagation. From the standpoint of ground-system resistance alone, however, conditions beyond $\lambda/2$ from the antenna are relatively unimportant.

The depth of the soil of good conductivity is important because of the very deep penetration of earth currents at low frequencies. It is often economically impractical to use ground systems of the radial type longer than about 500 m, and this may be but a small portion of a wavelength. A considerable portion of the ground current returns to the antenna base under the ground system, tending to produce high current densities near the center of the system where power is fed in. A large radial ground system will collect currents from the field in space over it and will be guided back to the feed point in the wires. The field beyond the wires sets up ground currents which return through the soil. If long ground rods are attached to the ends of the wires, these will intercept ground currents near the surface where they have greatest magnitude, but the deeper currents will pass under the system and some of the current will be refracted into the wires from underneath. However the greater portion will concentrate in the soil as they approach the feed point. Therefore, special means

¹ See references 88, 89, 94.

must be taken to collect these currents returning from great depth in a manner to minimize current densities in the soil and thus reduce ground-system resistance.¹ A circle of long ground rods attached to the ground wires a short distance from the feed point is desirable, as is also a large pipe driven deep into the soil directly under it and connected into the ground system. The length and number of radial wires are quite empirical, depending upon frequency, soil characteristics, and economics.

Another method of ground-system design for v-l-f systems is to employ from one to several circles of four to six short star ground meshes, each of six to twelve 100-ft radial wires with long ground rods at their ends. From the center of each star a

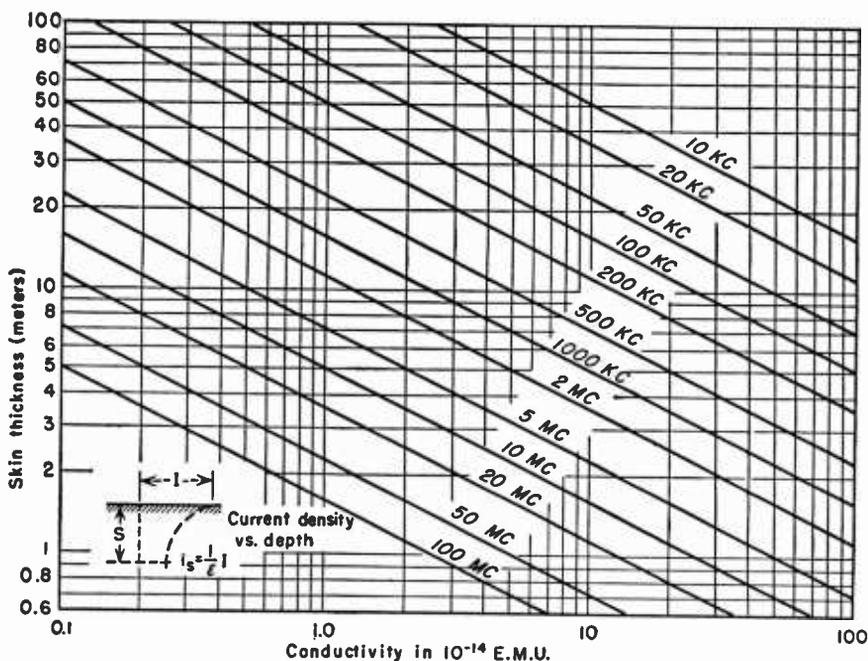


Fig. 8. Penetration of ground current. Skin thickness $S = (4\pi^2\alpha_{em}f\epsilon_0 \times 10^9)^{-1/2}$ cm.

conductor is brought back above ground, or insulated from ground, to the feed point or multiple-tuning point. When more than one circle of such stars are used, ground equalizer inductors are inserted in the shorter return wires so as to equalize the currents collected from each star. This technique reduces current densities in the soil by collecting ground currents in equal amounts from a large symmetrical area surrounding the antenna downlead.

Figure 8 shows depths of penetration of earth currents for various soil conductivities and frequencies. Approximately 90 per cent of the power loss in the soil occurs within the so-called skin thickness, which is that depth at which the current density has diminished to 37 per cent of that at the surface.

Capacitance grounds, or counterpoises, have certain merits which have long been recognized.² Instead of collecting ground currents by conduction from the soil, a large capacitance area composed of a system of wires a few feet above the earth is used. A virtually uniform electric field distribution exists under the counterpoise and causes an equally uniform distribution of earth currents beneath. The fringe effect increases its

¹ See reference 104.

² See reference 92.

effective area a few per cent. The area covered by a counterpoise should be as large as practicable to minimize ground losses. A small counterpoise located near the base of the downlead or tower, sometimes called a *ground screen*, may in some cases reduce local ground losses when used in conjunction with a buried-wire ground system.

16. Antennas for L-f Broadcasting. In those countries using the frequencies from 160 to 265 kc for broadcasting, there has been economic justification in employing antennas operating at or near the fundamental wavelength ($G = 90$ deg). This results in a relatively high radiation resistance and low reactance. In Europe there has been rather general adoption of a T-type single-wire antenna supported between two very high towers (as high as 250 m in some cases).

*Droitwich Antenna.*¹ A typical example which has been thoroughly described in the literature is that of the BBC station at Droitwich, England. A T antenna with a flattop 550 ft long and a downlead of 630 ft is supported between two towers 700 ft high and spaced 800 ft. The antenna consists of one wire. At the operating frequency of 200 kc, the impedance is $21 + j0$; at 90 kc, it is $17.5 - j41$; at 210 kc, it is $24 + j41$.

To correct for antenna selectivity, antenna coupling networks are applied sometimes which equalize the antenna impedance over a range of 4 to 6 kc each side of the carrier frequency. High single vertical radiators of large cross section, derived from the tower itself or by employing supported vertical wires around the tower to effect a large cross section, have the characteristics of lower intrinsic selectivity and high radiation efficiency.

17. High-tension Insulation for Wire Antennas. Very serious electromechanical problems are encountered in the insulation of large l-f antenna systems which must operate at high power inputs. Long spans, large conductors, heavy downleads, large total projected areas exposed to wind and sleet, and minimized sags so as to keep effective heights at a maximum, all contribute to enormous stresses at the insulator locations. Maximum electrical duty occurs simultaneously with maximum mechanical duty. Insulators must be of lowest possible capacitance, and the number used in parallel must be a minimum.

The most common type of insulator developed through the years consists of a glazed porcelain rod or tube cemented to end fittings having eyes for the rigging and permitting the attachment of rain shields when desired. The largest in use have 6 ft of clear porcelain and a diameter of 6 in. The porcelain is used in tension and is capable of supporting fair loads, provided no transverse impact is encountered. In tension, the accidental tap of the insulator with a tool will sometimes cause breakage. For years the problem of cement for the end fittings has been a major study. For high mechanical duty, insulators can be paralleled in a bridle or yoke. However, at best such insulators have indeterminate ultimate strengths, the high fragility is a great disadvantage, and they have limited life.

Another type of insulator has been developed for very high-tension antennas which has high mechanical strength and a determined ultimate working load even before actual breaking. This is the Austin safety-core insulator, which has taken a variety of forms for particular applications. It consists essentially of a cylindrical tube of glazed electrical porcelain with ends ground square, placed between end castings with smoothly turned inner surfaces to engage the porcelain ends with a gasket between them. Inside the porcelain tube and between the end castings is a high-tensile-strength phenolic link or Pregwood core which can be tightened as much as desired on one outside end. The assembly is placed in a hydraulic press which has a dynamometer to indicate the compression applied to the porcelain longitudinally by the end castings, and the core is tightened. Upon release from the press, the compression of the porcelain is balanced by tension in the core, while the porcelain remains in compression. The application of external pull first reduces compression in the porcelain and the core maintains constant tension until external forces are sufficient to pull the end pieces clear of the porcelain. This occurs when external pull exceeds the preloading applied during assembly. In compression, porcelain has its maximum mechanical strength and can withstand high transverse impacts as well. The shear

¹ See reference 90.

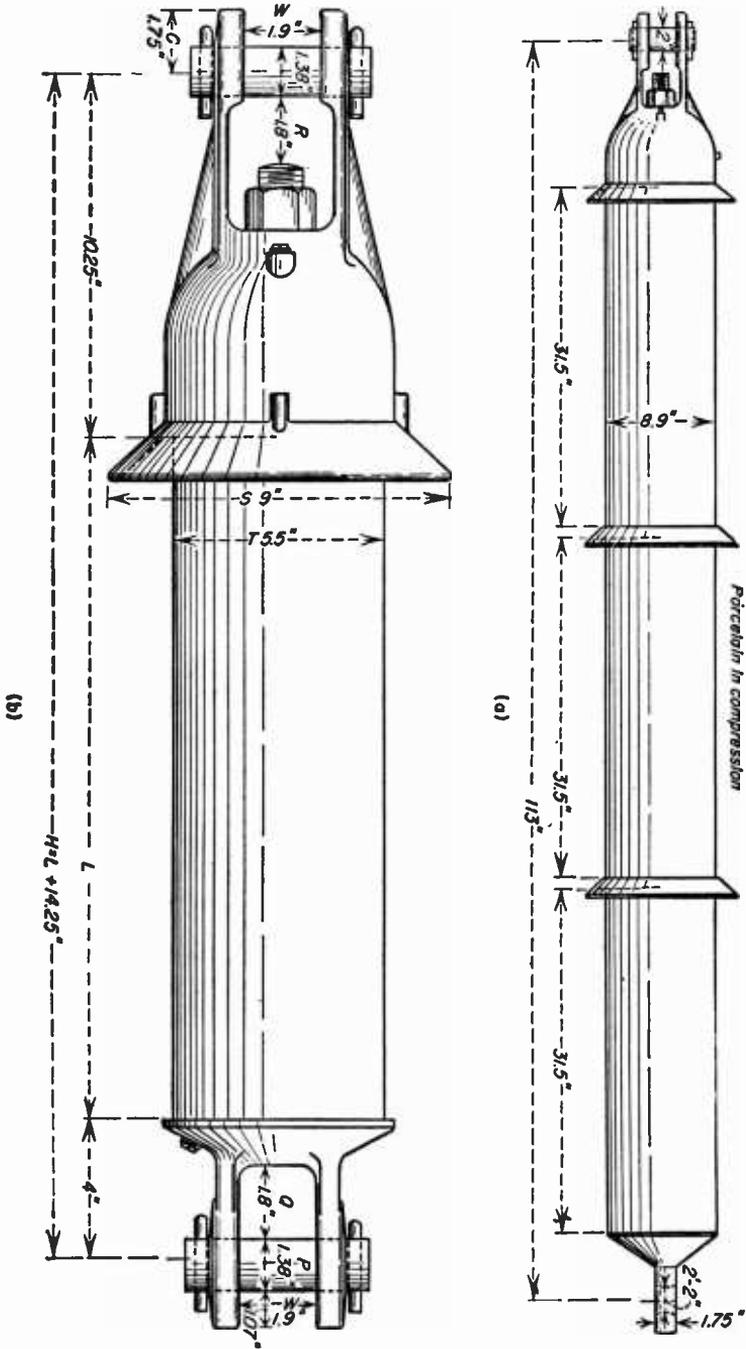


FIG. 9. (Left) Heavy-duty strain insulator, safety-core type, preloaded, oil-filled. (Right) Safety-core strain insulator, preloaded link-core type.

strength is also multiplied in compression. The assembled insulator is then filled with transil oil to remove all dead-air spaces which would be the points of origin of internal ionization.

The flux densities in the core and, therefore, the losses are very low as most of the electric flux passes through the porcelain tube when the diameter of the tube is large compared with the core cross section. Provisions are made to allow for oil expansion or contraction with temperature. Insulators of this type have been produced for r-f working potentials in excess of 200 kv under wet conditions, and working loads of the order of 80 tons tension. There are no cemented connections used, and the life is indefinite. Safety-core-type insulators for tower-base insulation and for guy insulation are available. The capacitances are slightly higher than the type described above but lower than other strain insulators of the conical compression type.

18. L-f Radio Range Antennas (200 to 400 KC). The use of low frequencies for radio guidance in aviation on a large scale in the form of 4-course radio ranges, the German Zonne system and radiophares generally has led to the vertical tower radiator for the elimination of night effect. Any emission of horizontally polarized energy from such systems causes large errors at night in courses and in direction finding. The Adcock type of radio range is now standard in the United States and Canada, uses four radiators (square array) or five (quincunx array) approximately 125 ft high spaced about 600 ft on a diagonal. These are usually base-insulated self-supporting towers.

There are several evolutionary versions of radiators and ground systems for the four-course ranges. With a ground system of fifteen 100-ft radial wires (buried) with end ground rods, for each radiator, the radiation efficiency is generally of the order of 8 to 10 per cent. Ground screens in the form of elevated metallic counterpoises about 30 ft square under the tower-base insulators are also used. They are series-tuned, and two outer radiators on a diagonal operate together to produce a figure-of-8 field-intensity pattern keyed systematically with the Morse A interlocked with N keying on the opposite diagonal pair. As power is switched between the two pairs of diagonal radiators, signals in the overlap region where the field intensities are equal produce a constant uninterrupted tone which provides a fixed track for guidance of aircraft. The ability of the ear to distinguish volume levels from the two alternated patterns give a course width of about 3 deg. Just at the place where the A or N keying becomes observable each side of the course is a narrow angle called the *twilight zone*, and aircraft ordinarily fly in the right-hand twilight zone. Between courses the A or N signals predominate.

When equal power is fed to the two halves of this system, reciprocal 90 deg courses are set up. If the power is lower in one pair, its radiation pattern is smaller than the other and reciprocal courses result but the courses are said to be *squeezed*. When both power and phase differences in the radiator currents are other than 180 deg in each pair, nonreciprocal courses are obtained. This is called *course bending*. These combinations are used to set range courses at angles best suited to airway routings in the four directions. The antenna system is energized through a goniometer, the rotation of which will turn the entire radiation pattern to the orientation desired.

The four-tower nonsimultaneous range switches the four towers into parallel in-phase feed when voice broadcasting is required, and during this interval the range courses are eliminated. The simultaneous range using five towers permits the four outer ones to transmit range courses continuously for guidance while the center tower radiates voice signals when required.

19. Whip Antennas. Whip antennas used on vehicles for transmission of h-f waves present much the same basic design problem as the electrically short l-f antenna. Radiation efficiency is intrinsically low, and care must be taken that it is no lower than necessary. In such systems tuning-coil loss and ground loss consume almost all of the transmitter power, and radiation efficiencies of the order of 0.5 to 2 per cent are common in the 2- to 8-Mc band. Base insulator leakage and capacitance should be as low as practicable. Anything that conditions of service permit which will decrease

the antenna reactance will contribute much to over-all improvement in efficiency. Some typical whip antenna data are shown in Figs. 10a and b.

20. Beverage (Wave) Antenna.¹ This type of antenna, one of the most effective l-f directive receiving systems to be used commercially, is a long transmission line. It is named after its inventor, H. H. Beverage, but is also called the *wave antenna*. A long open-wire transmission line, pointed in the direction of a downcoming wave, has a high degree of exposure to the horizontal component of the wave front, which induces

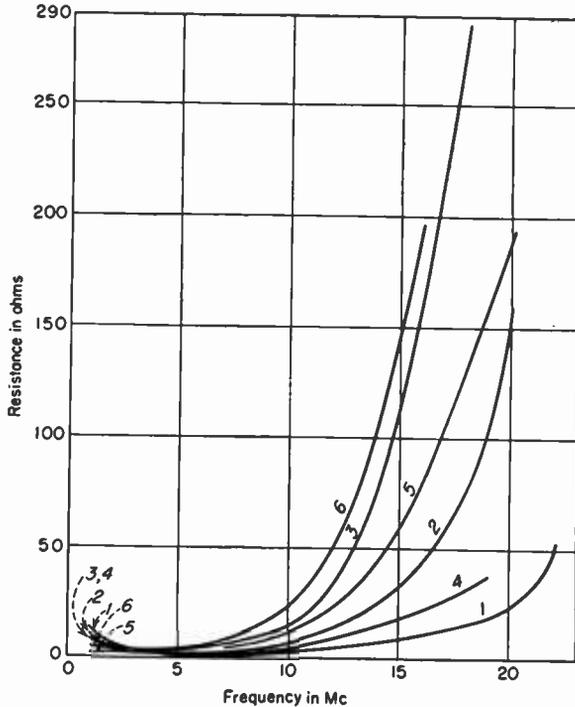


FIG. 10a. Measured resistance of various whip antennas mounted on an 8-cwt truck with truck chassis as ground: (1) 8-ft single rod; (2) 12-ft single rod; (3) 16-ft single rod; (4) 8-ft V aerial, 60 deg on common base insulator; (5) 12-ft V aerial, 60 deg on common base insulator; (6) 16-ft V aerial, 60 deg on common base insulator.

in the line a continuous series of emfs that are propagated along the wires in the form of a traveling wave. A wave front sets up a wave in the wire which starts at the distant extremity (in the direction of the arrival of the space wave) which is propagated toward the end where a receiver is situated. In addition, the entire wire receives energy from the downcoming wave, so that the effects are cumulative at the receiver, and a relatively large amount of energy is extracted from the space wave for energizing the receiver. The antenna functions only where there is an angular difference between the direction of the wire and the incidental direction of the space wave. This condition is suitably met in practice owing to natural conditions, since finite earth conductivity causes a wave traveling in space near the surface to be tilted forward at a considerable angle. Thus a long transmission line parallel to the surface of the ground

¹ See references 91 and 91.1.

has a workable inclination with respect to the wave front. This applies to vertical polarization.

The Beverage antenna has many useful forms which are specially adapted to long-wave reception, to short-wave reception, to bidirectional and unidirectional selectivity, for vertical and horizontal polarization, etc. A thorough treatment of these is impossible here, and detailed data must be obtained from the original and subsequent papers on the subject.

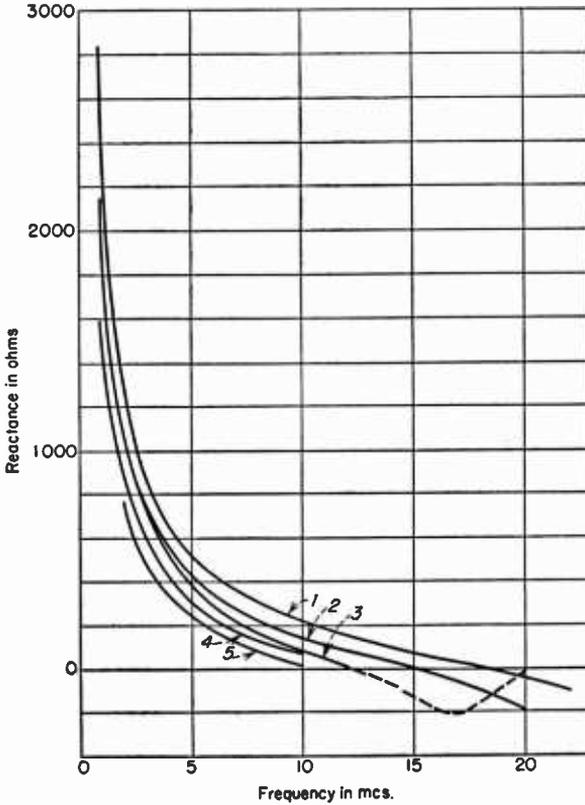


FIG. 10b. Measured reactance of whip antennas: (1) 8-ft single rod; (2) 12-ft single rod; (3) 16-ft single rod; (4) 12-ft V aerial, 60 deg on common base insulator; (5) 16-ft V aerial, 60 deg on common base insulator.

For long waves the antenna construction is very similar to ordinary open-wire telephone lines. The antennas may be located at a considerable distance from the station and coupled to the receivers by transmission lines. The Beverage antenna is directive in the line of its orientation and is made unidirectional by terminating the distant end in a resistance equal to the characteristic impedance of the line. Thus energy collected from a wave in the backward direction is completely dissipated without producing any influence in the receiver. Directivity may be sharpened by using two or more antennas in an array. This has been done in the Houlton system shown in Fig. 11 which is used for transatlantic telephone reception on long waves. One of the several forms of antenna used in this application is that which couples the receiver to the end of the antenna that is nearest the transmitting station. A two-wire line

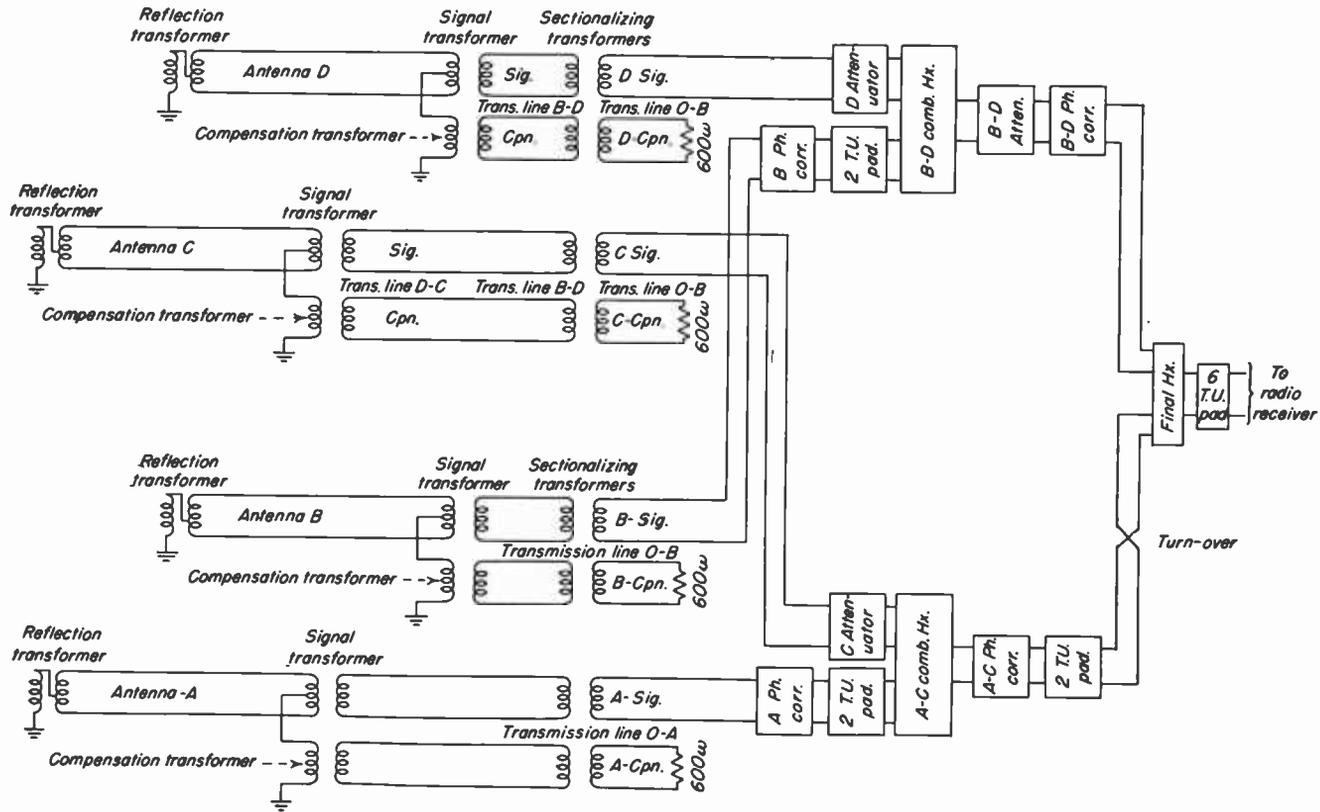


FIG. 11. Houlton antenna system for receiving transatlantic long-wave signals.

is used to achieve this in the following manner: Waves arriving from the preferred direction act upon the two wires in parallel to ground, and the induced wave of energy in the wire travels to the distant end where it encounters a reactive network called a *reflection transformer*. This device reverses the phase of the wave in one of the wires and reflects the energy from the end back to the receiver, the reflected wave of energy now traveling in the two wires balanced to ground.¹ The receiver coupling network terminates the line and absorbs all the wave energy in actuating the receiver. A wave entering the system from the reverse direction travels along the two wires in parallel against ground, produces no potential difference across the balanced termination, and therefore has no influence on the receiver. Instead, the circuit to ground is terminated in the characteristic impedance of the parallel-grounded system, and the unwanted wave is completely dissipated in a resistor.

In its very simplest form the Beverage antenna is a single straight horizontal wire a few feet above grade level, the length being anywhere from one to several wavelengths. The characteristic impedance of this wire unbalanced to ground is roughly calculable by using the image as the second conductor in a parallel-wire system. The receiver is coupled in at one end of the line, and the other end is terminated in a resistance equal to the characteristic impedance. Stable ground systems are necessary at both ends.

MEDIUM-FREQUENCY BROADCAST ANTENNAS (550 TO 1,600 KC)

When broadcasting began, the antennas used were smaller versions of the types previously used for lower frequencies. When the possibilities for getting higher field strengths along the ground by using high vertical radiators became known,² and greater knowledge of wave propagation in this band was obtained, a period of antenna development began which led to definite optimum designs,³ which are physically and economically realizable. While studies of aerial structures were being made, other detailed study was directed to ground systems, and again some definite practical optimum designs resulted.⁴ This work placed broadcast-antenna engineering on a firm foundation, which, together with the vast information on wave propagation by ground-wave and ionosphere reflections, has permitted an intense utilization of this band of frequencies for broadcasting services. This paved the way to the use of directive antenna systems which further increased the utilization of the limited number of channels for an ever larger number of stations and increased power.

Because of a number of fortunate natural circumstances, broadcast-antenna engineering is far simpler than that for most other frequency bands. For instance, a substantially perfect ground can be built under the antenna, permitting high radiation efficiency; within realizable heights of structures, the vertical radiation pattern can be adjusted for maximizing ground wave and minimizing sky wave, thereby decreasing the annular zone of overlap of the two and giving a high degree of effectiveness to both types of propagation by concentration of radiated power at angles close to the horizon. With modern data, broadcast-antenna performance can be predicted and measured with a remarkable degree of precision.

21. Nondirective Broadcast Antennas. Vertical polarization is required for broadcasting in this band, and vertical radiators are required to produce it. One steel tower used as a radiator costs less than two towers supporting an equivalent wire antenna and eliminates field distortion caused by currents induced into the supporting towers. Hence present-day broadcast antennas are almost exclusively the tower-radiator type, either guyed or self-supporting. Structures having uniform or nearly uniform cross section are necessary to assure a predictable vertical radiation pattern from a predictable current distribution. Top loading, sectionalizing, and combinations of these two techniques are occasionally employed to produce an optimum vertical direc-

¹ A similar effect is obtained by grounding one wire and leaving the other open-circuited.

² See references 7 and 8.

³ See references 101, 103, and 121.

⁴ See reference 104.

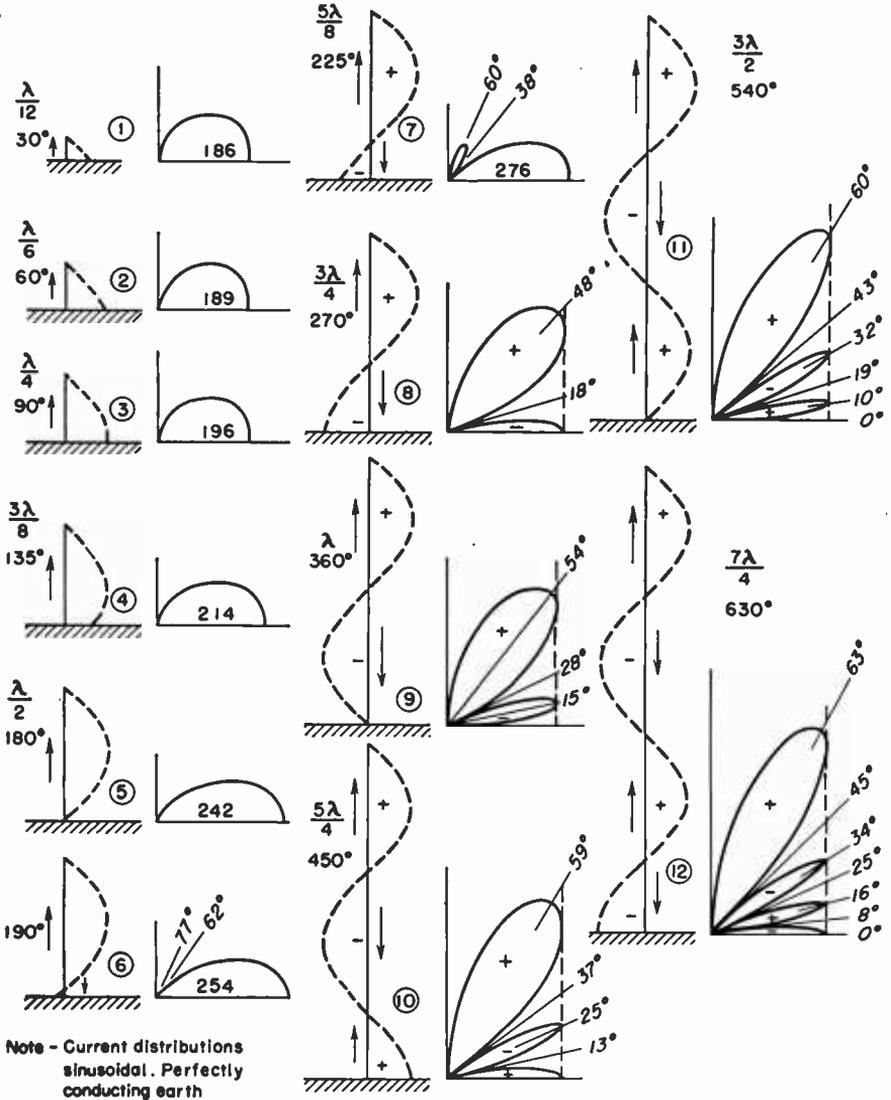


FIG. 12. Vertical field-intensity patterns for vertical antennas.

tivity with limited tower heights. Tower radiators have further advantages derived from their relatively large cross section, which provides a lower characteristic impedance than a wire antenna, with lower potential gradients for the same power, lower reactance, and greater band-width capability for a given impedance variation.

22. Current Distributions and Vertical Radiation Patterns. As the electrical length of a uniform vertical conductor is increased, a larger and larger portion of a complete near-sine wave of current is distributed on it. In space, each element of

length produces a field whose intensity is directly proportional to its current amplitude. The integrated fields from all parts of the antenna, together with those reflected from the ground, interfere in varying degrees from complete addition to complete cancellation, for different vertical angles. The over-all result of this interference in the half space above ground is called the *vertical radiation pattern*. For typical broadcast applications it is given in terms of relative or actual field intensities. Antenna lengths which slightly exceed $\lambda/2$ of sinusoidally distributed current produce a vertical radiation pattern flattened toward the horizon, and this vertical directivity causes a gain in field intensity along the ground. Starting from very short length, the pattern of which is a semicircle through any vertical plane through the axis of the antenna, the vertical pattern changes very slowly with height until it exceeds $\lambda/4$. Between $\lambda/4$ and $\lambda/2$, the pattern flattens considerably. Greatest field intensity along the ground

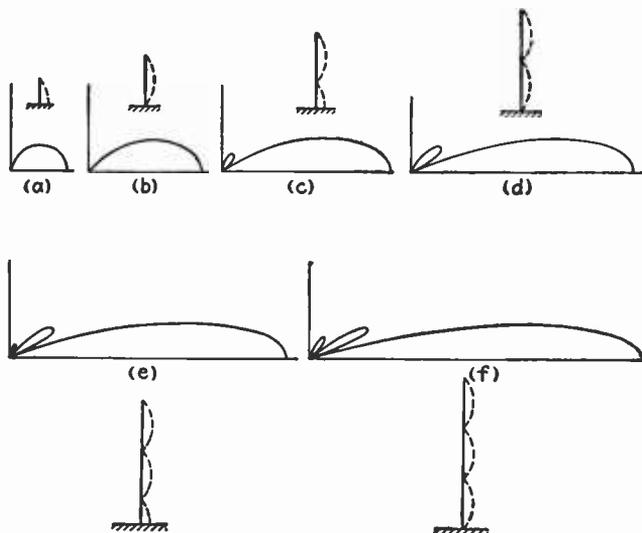


FIG. 13. Polar patterns of the field-strength distribution for vertical antenna over perfect earth when the currents in successive dipole sections are cophased.

occurs with a height of $5\lambda/8$ accompanied by a secondary high-angle lobe which is undesirable for its effect on short-distance sky-wave interference with the ground wave at night.¹ Beyond $5\lambda/8$, the ground lobe rapidly contracts, and energy is transferred to the high-angle lobe and thus becomes unsuitable for broadcasting.² Thus, from the standpoint of horizontal gain, $5\lambda/8$ represents a maximum condition for natural current distributions. If special means are taken to prevent a reversal of currents below the upper half wavelength of the radiator, further horizontal gain can be obtained with increasing antenna height. This scheme, due to Franklin, is exemplified in Fig. 13.

The dependence of the vertical radiation patterns on current distribution and antenna height are shown in Figs. 12 and 13. In Fig. 12 diagrams 1, 2, 3 conform to the polar equation

$$f(\theta) = m \sin \theta$$

Diagram 4 has the equation

$$f(\theta) = \frac{m \cos (90^\circ \cos \theta)}{\sin \theta}$$

¹ See reference 8.

² See references 56 and 101.

All diagrams except 7, 13, 14 conform to the more general equation derived from a sinusoidal current distribution in a vertical antenna whose lower end is at ground level, and for total height G deg.

$$f(\theta) = \frac{m \cos (G \cos \theta) - \cos G}{\sin \theta (1 - \cos G)}$$

where θ is the angle from the antenna axis and G the electrical height of the antenna.

Values of the coefficient m in millivolts per meter over perfectly conducting earth at 1 mile with 1,000 watts radiated are shown on the various diagrams. It is important to note that for diagrams 1 to 4 (a range of heights most frequently encountered in practice) the value of m increases only 5 per cent in going from very small height to a full $\lambda/4$; in going to $3\lambda/8$ it increases only 15 per cent over that for a very short radiator.¹ These facts have an important bearing on antenna economics but are contingent upon simulating virtually perfectly conducting ground by using a suitable ground system.

Figure 14 is from the FCC "Rules of Good Engineering Practice for Standard Broadcast Stations."²

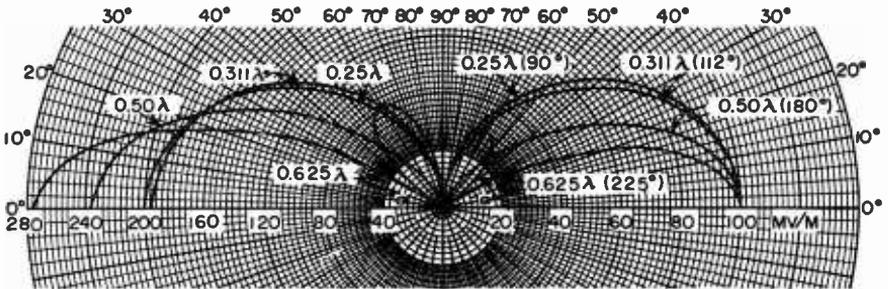


Fig. 14. Vertical radiation patterns for different heights of vertical antennas with sinusoidal current distribution.

23. Obstruction Lighting. When an antenna system is designed, the regulations regarding the lighting of high structures that constitute obstruction to aviation should be carefully investigated and the necessary details anticipated in the design. Regulations for lighting and painting vary in different countries and with remoteness from airports and established airways. Lighting circuits often must include flashing beacons and spare lights with automatic transfer in case of failure. In some types of antennas the design is seriously affected by the needs of obstruction lighting. Since the antenna system must function correctly with the lighting circuits, all operating measurements must be made accordingly.

When towers are used as radiators, lighting power must either pass through all the coupling and tuning devices, or by-pass them through double-conductor chokes of very high r-f impedance, or through special toroidal transformers with air spacing between windings, such as the Austin tower lighting transformers. The latter has the characteristic of low capacitance and very high flashover potential between windings. Several versions of the latter are available, including one built into a safety core base insulator and immersed in transil oil and capable of withstanding extremely high r-f potentials.

24. Bonding between Sections of Tower Radiators. Special care is required during construction to ensure good electrical contact between tower sections. Good electrical bonding may be required across the joints between vertical sections. To aid in providing good electrical conductivity throughout the length of a tower, galvanizing of the

¹ See reference 104.

² See reference 118.

steel is desirable. Towers used as supports do not require galvanizing except in situations where painting is not required for aviation obstruction marking, in which case the spelter provides the necessary protection against corrosion. Where painting is required, the paint will serve as adequate protection when renewed at intervals specified to keep the colors bright. For a tower radiator the paint goes on over the galvanizing.

25. Ground Systems.¹ A ground system consisting of 120 buried radial wires (depth of burial sufficient for protection only) a $\lambda/2$ long, provides a virtually perfect grounding terminal. This length is sufficient to collect earth currents due to the antenna field above ground over the top of the ground wires, and only a negligibly small amount of earth current is collected from beyond the ground wires. Such a system provides a radiation efficiency in typical cases of the order of 95 to 98 per cent in terms of short-distance field intensities realized. A further advantage of this system is with antifading antennas, where the high-angle radiation pattern retains its theoretical form due to the very small reflection losses to waves reflected from the area of the ground system, which functions as a perfectly reflecting surface.

The degree of compromise in the length and number of radial wires in the ground system depends upon the performance requirements. In any event, the number of radials should not be reduced even though in some cases the length is reduced to conserve land. To meet a specified efficiency of 150 mv per m at 1 mile with 1 kw, using radiators of height greater than $\lambda/8$, the ground radials could be reduced in length to 0.25λ . For higher efficiencies, longer radials are necessary.

In radial ground systems as large as these, the electric flux from the antenna terminates on the ground wires from above, and the current densities under the wires are very low. For shorter wires, where earth currents return to the system from beyond the wires, ground rods at the ends may be justifiable, as with l-f systems. There is no need for circumferential bonding of the radials.

Directive antennas employing two or more radiators require a separate radial ground system for each up to the points where the wires of one system cross those of another, at which points they can be bonded and the projections eliminated.

26. Directive Broadcast Antennas. The allocation standards of the North American Regional Broadcasting Agreement pose some difficult design problems for new broadcast stations, or those seeking power increases.² Most of these require directive antennas of varying complexity. Under the allocation rules, the limiting field intensities at particular azimuth and altitude angles, as determined by geography and wave propagation, are calculated, and a directive antenna capable of performing within such limits must be found. A great deal of ingenuity and computational labor is involved in finding the specifications for an acceptable array. To aid in this task, some calculating devices have been developed to shorten the work of cut-and-try approach for three radiators. Very recently, however, the RCA Antennalyzer has helped solve pattern problems for two to five radiators in any azimuth, spacing, current amplitude, and phase relationship. With five radiators there are 16 independent variables that must be determined in strictly correct combinations. To do this, the Antennalyzer presents a horizontal pattern in either polar or rectangular coordinates on a cathode-ray oscillograph screen in correct azimuth. The dials are adjusted until a pattern is obtained that meets the limits necessary, and the specifications for the array are read directly from the dials. The pattern can also be made in rms form in a simple operation. Beside its value as an array synthesizer, this instrument is valuable in studying pattern stability problems, since tolerances in any variable can be quickly observed.

Without such an instrument, one proceeds the hard way to find an array giving a pattern meeting prescribed limits of protection and stability with the desired power.

27. Two-element Arrays. A wide range of broadcast directive patterns can be had using two vertical radiators by varying the tower-line azimuth, the spacing, the

¹ See references 104 and 16.

² See reference 118.

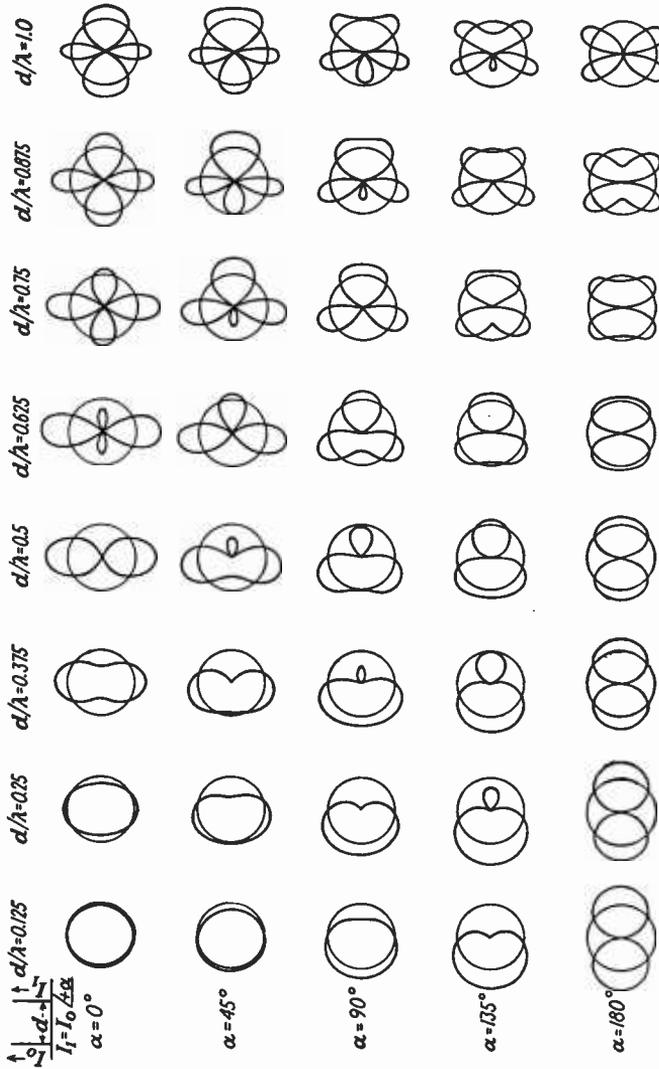


Fig. 15. Horizontal patterns for two radiators fed with equal magnitude currents taking mutual impedance into account. Circles indicate relative fields from one radiator alone with same power input.

current ratio, and the phase difference.¹ Such patterns are symmetrical about the vertical plane through the radiators. A two-radiator directive system is the simplest that can be applied to directive broadcasting. Its pattern can be calculated from the equation

$$F_{\theta, \phi} = f(\theta)(1 + k_1 e^{i\psi_1})$$

The legend for this equation is given later. For the case of $k = 1.0$ (equal radiator currents), the horizontal field-intensity patterns over a range of values of spacing and phase differences are presented in Fig. 15, which include the effects of mutual imped-

¹ See references 26, 102, 79, and 130.

ance. The reference circle in each diagram is therefore the relative field intensity for one radiator with the same power input.

28. Radiation Pattern for Generalized Multielement Array. The radiation pattern for an array of n identical vertical radiators whose ends are normal to a common plane (such as ground), as observed on the surface of an enclosing hemisphere whose radius is very large with respect to the greatest dimension of the array is given in vector form as follows:

$$F_{\theta, \phi} = f(\theta)(1 + k_1 e^{i\psi_1} + k_2 e^{i\psi_2} + \dots + k_{n-1} e^{i\psi_{n-1}})$$

where $F_{\theta, \phi}$ = relative field-intensity pattern in azimuth ϕ and altitude angle $(90 - \theta)$

$f(\theta)$ = relative vertical field-intensity pattern for 1 radiator

$k_m = I_m/I_0$, and may be greater or less than unity

$e^{i\psi_m}$ = unit vector at an angle ψ_m with respect to reference vector

ψ_m = phase angle due to space and time differences and

$$\psi_m = S_m \cos(\phi - \phi_m) \sin \theta + \alpha_m$$

where S_m = spacing in electrical degrees of radiator m with respect to radiator O

α_m = phase difference between I_m and I_0 ; ranges through values of ± 180 deg

ϕ = angle measured clockwise from reference azimuth, usually true north

ϕ_m = azimuth of line through radiator m and radiator O and ranges clockwise through 360 deg

For the horizontal pattern only, $\theta = 90$ and $\sin \theta = 1.00$. For the vertical pattern through any azimuth angle ϕ , $\cos(\phi - \phi_m) = \text{constant}$.

Each term in this equation represents the vector contribution of one of the radiators. When all radiators are located on a common line, the radiation pattern is symmetrical with respect to the vertical plane through this line.

The rms value of a horizontal pattern can be obtained by measuring the area of the polar plot and constructing a circle of the same area.¹ If done in rectangular coordinates, the relative field intensities at each azimuth must be squared, then the area under the squared curve measured. The rectangle that has this same area on the same base then has its altitude squared to give the altitude of the rms field intensity for a single radiator with identical power input. From known performance of single radiators in terms of field intensity, power, and efficiency, the field-intensity scale in absolute values can be determined for the directional pattern.

29. Array Design. The physical realization of a calculated directive antenna involves great skill in design and adjustment, and great care to accuracy of details. Through radiation couplings, any change in spacing, azimuth line for the radiators, current ratio, and phase difference of any radiator current reacts upon every part of the entire radiator and feeder system. Even with simple two-radiator systems, cut-and-try methods of adjustment are virtually hopeless. It is necessary therefore to predict accurately the conditions which must exist throughout the system when working as desired, and by construction and careful measurement, to provide those conditions precisely. This naturally becomes more difficult as the number of radiators increases. In symmetrical systems there is a helpful simplification of work by dealing with radiators in pairs. The use of identical radiators of accurately known self-impedance and current distribution is particularly important as a starting point. For close-spaced radiators of the antifading type, the current distribution is modified by mutual impedances which in turn modifies the expected feed-point impedances, and introduces empirical complications. Care must be exercised that guys on radiators do not measurably deform the basic circularity of pattern for each radiator. Ground systems must present identical resistances in each radiator.

In designing a directive antenna system, the feed-point impedance of each radiator must be calculated for conditions which will produce the desired radiation pattern. This is done as follows:

Let $Z_{00} = Z_{11} = \dots = Z_{mm}$ (from measurement of one prototype radiator at and

¹ See reference 143.

about the working frequency, including base insulators, tower lighting devices, phase monitoring circuits, and any other attachments actually employed in final working status). However, identity of radiator impedances is not always assured by identical construction.

The transforming effect of all attachments to the radiator should be known as accurately as possible to be able to correct the mutual-impedance effects as transferred to the feed points, since the available mutual-impedance data are based on ideal conditions without such transforming effects present.

Measured values of Z_{00} (reference radiator) are then converted to their value *inside* the transforming devices. This then provides the idealized values of self-impedance.

Mutual impedances between all combinations of the radiators are then calculated or read from curves,¹ such as Fig. 16.

Simultaneous equations are then set up including the vector current relationships required in the various radiators, in the idealized location beyond the transforming effects:

$$\begin{aligned} E_0 &= I_0 Z_{00} + I_1 Z_{01} + I_2 Z_{02} + \cdots + I_m Z_{0m} \\ E_1 &= I_0 Z_{10} + I_1 Z_{11} + I_2 Z_{12} + \cdots + I_m Z_{1m} \\ E_2 &= I_0 Z_{20} + I_1 Z_{21} + I_2 Z_{22} + \cdots + I_m Z_{2m} \\ &\dots\dots\dots \\ &\dots\dots\dots \\ E_m &= I_0 Z_{m0} + I_1 Z_{m1} + I_2 Z_{m2} + \cdots + I_m Z_{mm} \end{aligned}$$

The vector ratio of E_0/I_0 , E_1/I_1 , E_2/I_2 , etc., for each radiator then gives the complex impedance (idealized) for each radiator. These impedances are then transformed

back to the actual feed-point impedances (taking into account any phase shifts introduced by the transformation) for the several radiators.

Then the power input to each radiator must be calculated for the currents used in the previous calculation (based at this stage on current ratios) and the division of power among the several radiators determined. With power distribution known and also the feed-point impedances, the feed-system and the coupling-phasing networks can be calculated for each radiator and for the power-dividing networks. The designer has a wide latitude of choice in the arrangement of feeders, the type and characteristic impedance to use, and in the location of the power-dividing networks.

Owing to the fact that mutual-impedance vectors can lie in any of the four quadrants,² it is possible to find that, in some circumstances, the input impedance to a radiator has a negative

resistance. This of course means that instead of feeding power into this radiator, a certain amount of power is delivered from the radiator into the feeder system. This is negative power from the standpoint of determining power division, and the feeder line must be terminated at the opposite end. Wherever possible to do so, it is desirable to avoid negative resistance inputs to radiators in the interests of system simplicity. However, in complicated arrays this may be unavoidable.

¹ See references 19, 44, 45, and 102.

² See references 102, 44, and 45.

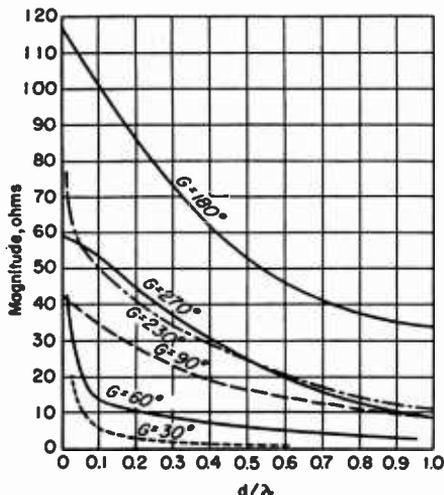


FIG. 16a. Mutual impedance between antennas.

It is customary to terminate all feeders so as to suppress standing waves. Then the phase lag in a feeder is equal to its electrical length, which in turn depends upon the velocity of propagation in the lines used. If open-wire lines are used, the velocity is the same as that in free space. Enclosed lines always have velocity reduced in some degree, and this must be accurately known in order to use the electrical length of the line for the phasing calculations.

It is not necessary to suppress standing waves in the feeders unless they are so high as to cause appreciable power loss or high voltages. When not terminated, the line acts as a transformer, the input impedance is generally complex, and the phase lag for the line current is not the same as the electrical length of the line (see Fig. 17).

In a directive antenna it is the current that must have proper phase and amplitude. Thus, in the design of the feeder and coupling system, the current phases are to be controlled. Starting with the transmitter, the total phase difference to each radiator, the power dividing networks, feeders, and coupling networks, must be such as to

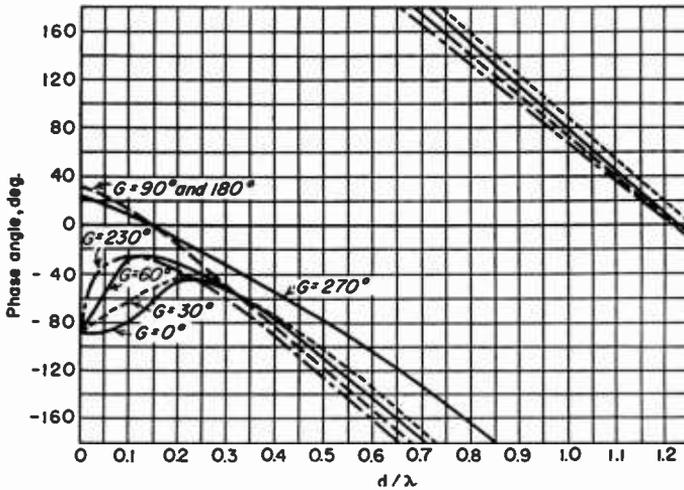


FIG. 16b. Phase-angle effect of coupled antennas.

bring to each radiator its correct relative phase and amplitude. Most of the networks will have to make a specified impedance match with a specified phase shift, and networks must be individually synthesized on that basis.

When the system has been fully calculated and constructed ready for adjustment, the individual elements of each network are adjusted to calculated values at the working frequency with suitable measuring instruments.¹ Inductance adjustments must include external leads. After this operation, the network can be further verified by connecting an impedance across its output terminals equal to the anticipated working load and having its input impedance measured, trimming adjustments until correct input impedance is noted. Such trimming should involve only very small changes or otherwise the phase-shift characteristics of the network may be incorrect even though the impedance transformation is correct.

When connected completely, the entire system can be energized with power, and the various currents in the system checked for amplitudes as a first test. If correctly calculated and measured, very nearly correct relative values should result immediately. Next, phases are checked by a phasemeter, or by the ultimate test, the measurement of the pattern. A first measurement of field intensity is made to determine that the

¹ See reference 102.

nulls and minimums occur at correct azimuth angles. Careful circuit trimming ensues until the ammeters (previously checked for correct calibration) show correct relative values and the pattern nulls occur at correct angles.

Final test is to measure carefully the entire horizontal pattern at points of small enough azimuthal difference to provide full data in relation to the complexity of the pattern. Measurements must be made on each angle to permit calculation of the unattenuated field intensity at some constant distance. For large arrays this may be the 2- or 3-mile circle, so that the pattern is fully formed. If the pattern is found to have the correct shape and orientation, power adjustments are then made for rated input, upon which basis the proof of performance is made.

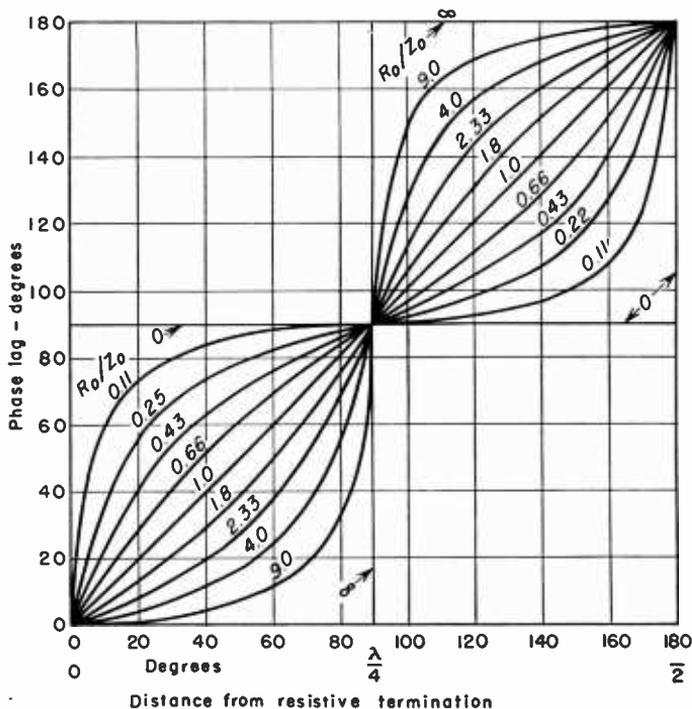


FIG. 17. Phase lag vs. line length for dissipationless line with resistive termination for various ratios of load resistance to characteristic impedance R_0/Z_0 .

30. Stability of Directive Antennas. Stability includes the effects of changes of the design parameters as well as the frequency. During modulation, side bands may be observed in a deep null even though the carrier is properly suppressed. The greater the degree of suppression of radiation in null directions, the larger the number of radiators employed, the greater the band width due to modulation, the greater the problem of stability. Many theoretical designs become impractical when viewed from the standpoint of stability. All system connections must be thoroughly secure, and network-connecting leads held rigidly in their tested positions. Effects of corrosion, weather, deterioration, heating, etc., must be carefully anticipated to assure stability of adjustment for long periods. Networks should be placed in tamperproof locked boxes. With such precautions, directive systems maintain excellent long-term stability.

31. Antennas on Ships. Ships of all sizes, with their masts, rigging, derricks, etc., severely circumscribe the antenna designs which can be used. As time goes on, more and more antennas are needed for new radio services on ships. There is usually the "main" antenna for l-f transmission and reception, one or more h-f antennas, a sense antenna for the direction finder, an antenna for the ship's broadcast receiver, and occasionally others for special uses. Their disposition depends largely upon the architecture of the ship and the location of the various facilities and operating rooms. They are located wherever space can be found.

The main antenna is usually a single-wire inverted-L, or T antenna suspended between the main masts. Being electrically short at the lowest working frequencies, there is a rather high potential on the system when used with medium- and even low-power transmitters. Therefore, the insulation must be electrically and mechanically robust throughout its length and through the deck bushing to the transmitter. This antenna is often used also for h-f service, in which case the system is electrically very long and provides a many-lobed pattern with vertical and horizontal polarizations. The random radiation patterns obtained provide the necessary functions for communication. The same is essentially true for the single-wire antennas installed for h-f duty only.

The sense antenna for use in conjunction with the direction-finder loop antenna is as nearly vertical as can be and terminates at the d-f location.

As a consequence of typical shipboard circumstances, there is little or no electrical designing possible in the sense in which it is applied to types of land services. There are important mechanical design considerations to be met to withstand the severe weather encountered at sea.

32. Antennas on Aircraft. Only a few elementary remarks will be made on this complicated subject, and these apply to the simplest applications on transport aircraft. The characteristics of aircraft antennas are largely empirical, depending upon the frequency, the size and shape of the aircraft, and the available locations. With present commercial flying speeds, fixed antennas are the rule and so disposed as to offer a minimum drag during flight. Thus antennas are generally oriented parallel to the slip stream over the fuselage (for l-f and h-f services) from a short mast over the cockpit to the tail structure, and under the fuselage for the 75-Mc marker antennas. Vibrational stress is severe, and precautions must be observed in choice of material and method of rigging to sustain typical conditions. At the frequencies where the antenna has high reactance and low resistance, a few watts of antenna power produce very high potentials, which, with high altitude, easily gives rise to corona and flash-over. Corona formation in high-velocity air occurs at higher potentials than in still air. Care is required to avoid sharp points which accentuate discharge, and the conductors should be as large as possible. Insulated antenna wire raises the corona potential and also reduces precipitation static.

Direction-finder loops are sometimes fully exposed in low-speed ships, enclosed in streamlined envelopes for medium-speed ships, and located inside the nose of the ship behind a dielectric envelope in high-speed ships. To reduce the size of exterior loops and thus decrease drag for exterior loops, powdered-iron cores can be used. When two loops are used with automatic direction finders, exterior mounting is usually necessary.

For instrument landing, both localizer and glide-path antennas are usually located somewhere near the cockpit or the radio room. For v-h-f operation these are presently small horizontal single-turn fixed loops with balanced outputs to the equipment. Vertically polarized v-h-f communication equipment often uses a short whip antenna. U antennas can also be applied to v-h-f reception of horizontal polarization. The art of v-h-f and u-h-f antenna design will certainly evolve rapidly in the future and lead to certain forms of suppressed antennas built into the skin of the ship.

Radio altimeters employ a pair of electric dipoles, of the type shown in Fig. 36b, one located under each wing with the dipole parallel to the slip stream. Careful streamlining of the antenna is required for this application, and a spacing insulator is used at the center for mechanical rigidity.

Trailing wire antennas are impractical for high-speed aircraft unless they are very short. A trailing wire v-h-f antenna, with coaxial transmission line and a coaxial dipole antenna at its end for the antenna and feeder system from the rear of the ship, may be used in some circumstances. For private aircraft, a prevailing type of h-f trailing antenna passes from the antenna reel through a guide at the top of the tail structure, and is held out in the slip stream by a small wind sock at its end. When reeled in, this antenna is secure and obscure.

For many important radio applications in aircraft, the radiation pattern for the antenna must be of prescribed shapes, and great difficulty is encountered in achieving the proper electrical, mechanical, and aerodynamic characteristics.

ANTENNAS FOR H-F APPLICATIONS (3 TO 30 MC)

33. Straight Horizontal Half-wave Dipole. A straight conductor parallel to earth, of the order of one-half wavelength long, suitably excited, is one of the most elementary and useful h-f radiating systems. When exactly one-half wavelength long, and in free space, a thin wire antenna has a center-point impedance of approximately $73 + j30$ ohms. When placed parallel to and within about 5 wavelengths of ground, mutual impedance with its image causes the impedance to vary appreciably with its height. Figure 18 shows how the radiation resistance varies with height for both horizontal and vertical half-wave dipoles.¹

The electric field around a $\lambda/2$ dipole resembles that of the magnetic field of force around a short bar magnet. When in free space, the field intensity at constant distance follows the relation

$$F_{\theta} = \frac{\cos(90^{\circ} \cos \theta)}{\sin \theta} \quad (\theta \text{ measured from wire})$$

from which it can be seen that there is zero field intensity in the direction of the dipole and maximum everywhere at right angles to it. It is therefore a directive radiating system. When used in practice over ground, some endwise radiation occurs because of reflections from ground and from the ionosphere, and in this direction the field is vertically polarized. Normal to the wire the field is horizontally polarized. In intermediate directions there are components of both.

In h-f applications, the most important characteristic of such an antenna is its vertical-plane pattern normal to the wire. Over perfectly conducting earth, this has the equation

$$F_{\phi} = \cos(H \sin \phi \pm 90^{\circ})$$

When the complex reflection coefficient of the ground is taken into account, the equation becomes

$$F_{\phi} = 1 + K_{\epsilon}^{-j2H \sin \phi}$$

where F_{ϕ} = relative field intensity as a function of elevation angle ϕ

H = electrical height of antenna above ground

K = complex reflection coefficient for horizontal polarization and is a function of soil conductivity σ and inductivity ϵ , the angle of incidence ϕ and the frequency f , derived from the equation

$$K = \frac{\sin \phi - (\epsilon - j2\sigma/f - 1 + \sin^2 \phi)^{1/2}}{\sin \phi + (\epsilon - j2\sigma/f - 1 + \sin^2 \phi)^{1/2}}$$

The phase angle is always 180 deg in this case.

For perfectly conducting ground (hypothetical), $K = 1$. Figure 19 shows a series of these patterns for various heights up to $2\frac{1}{2}$ wavelengths. The relative polarity of the field in each lobe is indicated. Figure 20 charts the angles of the maximums and the nulls for these patterns. These angles are important for propagation purposes.

¹ See reference 19.

Figure 21 exhibits the optimum radiation angles for propagation via the various ionosphere heights as functions of distance for one-hop transmission.

Both transmitting and receiving antennas of this type, used over fixed circuits, should be at the same height and perpendicular to the direction of transmission.

34. Feed Systems. The straight half-wave dipole can be center-fed in series from a balanced feeder, or shunt feed (also called delta- and Y-feed).¹ In the former, there is a large mismatch between antenna and feeder impedances for the type of feeders commonly used so that means must be employed to match the line at some point

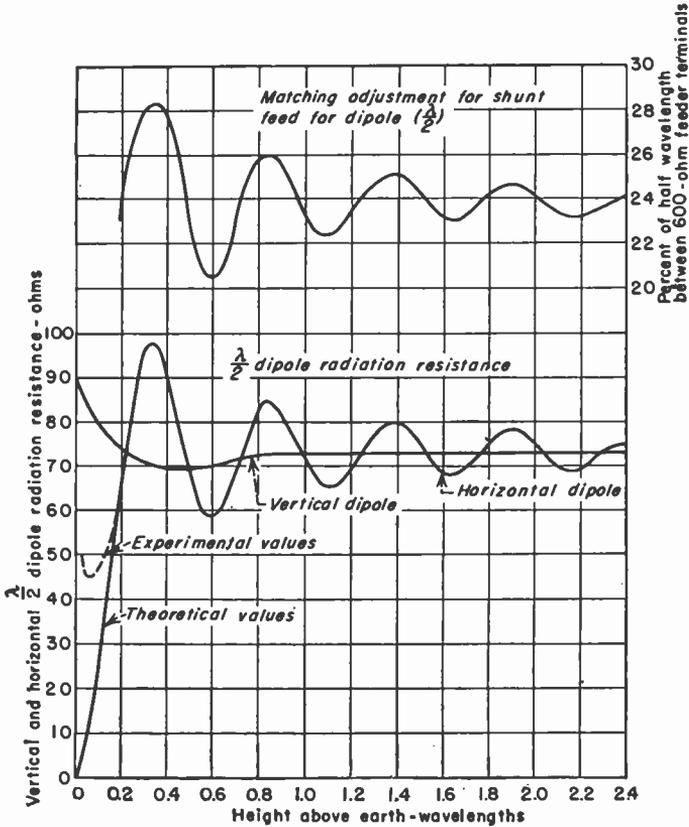


FIG. 18. Radiation resistance vs. height.

close to the antenna if highest efficiency is desired in long feeders. The shunt-feed system, when correctly performed, makes a satisfactory impedance match directly (but never perfectly, unless series capacitors are included to correct for the inductive effect of the enclosed portion of the antenna) for low standing-wave ratio on the feeder. Care must be taken to maintain exact symmetry of connection between line and antenna. Adjustments for this are shown in upper part of Fig. 18.

Feeders should always run normal to the antenna wire for as far as possible from the antenna, and skew relations avoided wherever possible. Skew relations cause radiation couplings between feeder and antenna which upset line balance and com-

¹ See reference 19.

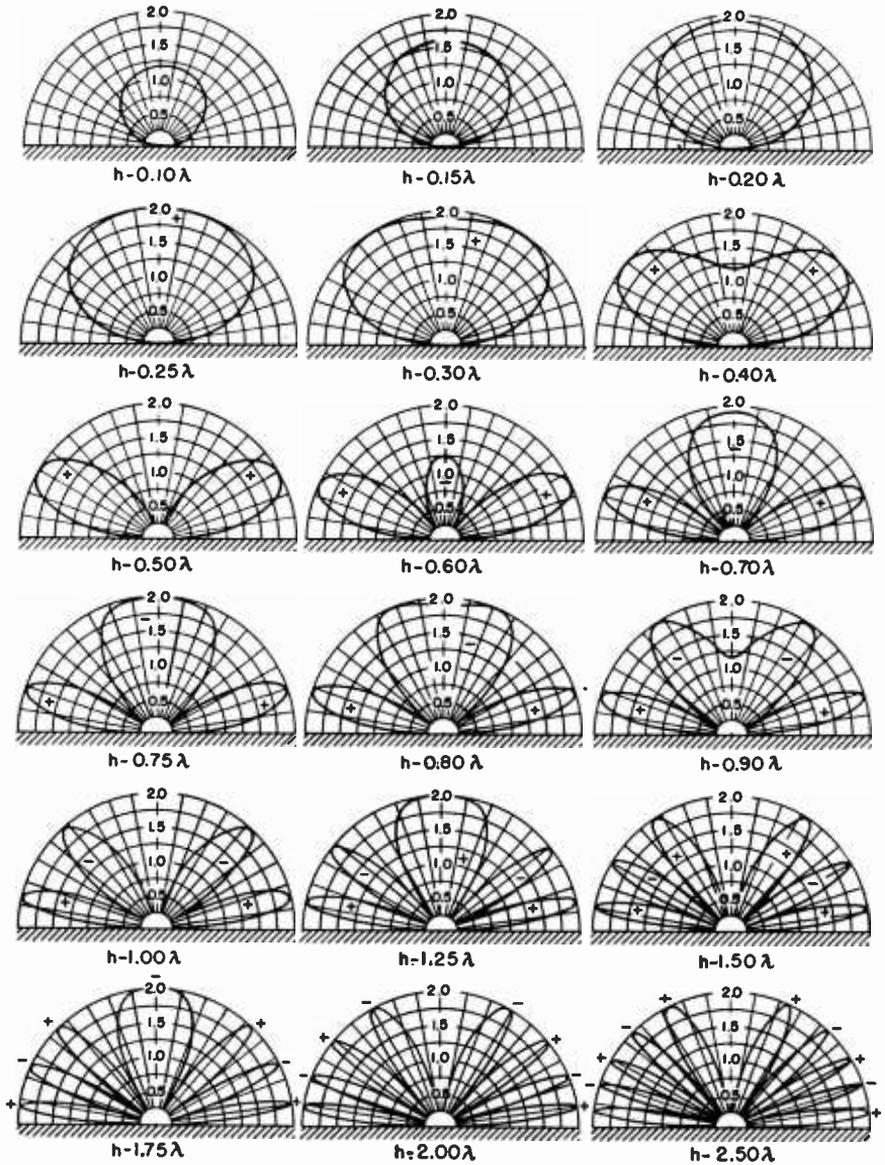


FIG. 19. Vertical polar diagrams for horizontal dipoles.

promise pattern and efficiency. This is quite general and should be observed in all antenna designing.

Instead of spreading the feeder to make an impedance match, the dipole can be shortened and a small inductance inserted between the junctions with the feeder. The latter is run at fixed separation directly to the antenna. A compromise match

can be made in this way, and the antenna is then known as a *shortened dipole*. Its advantage over the simpler arrangement preceding is nil, and the operation of adjusting is more difficult.

Another feed method, technically inferior but often used for its simplicity, is off-center feed, with a one-wire feeder, working against ground.¹ A position can be found on the antenna where standing waves on the feeder are quite small for one frequency.

Transformer coupling between feeder and antenna can also be used but there is no good reason to do so.

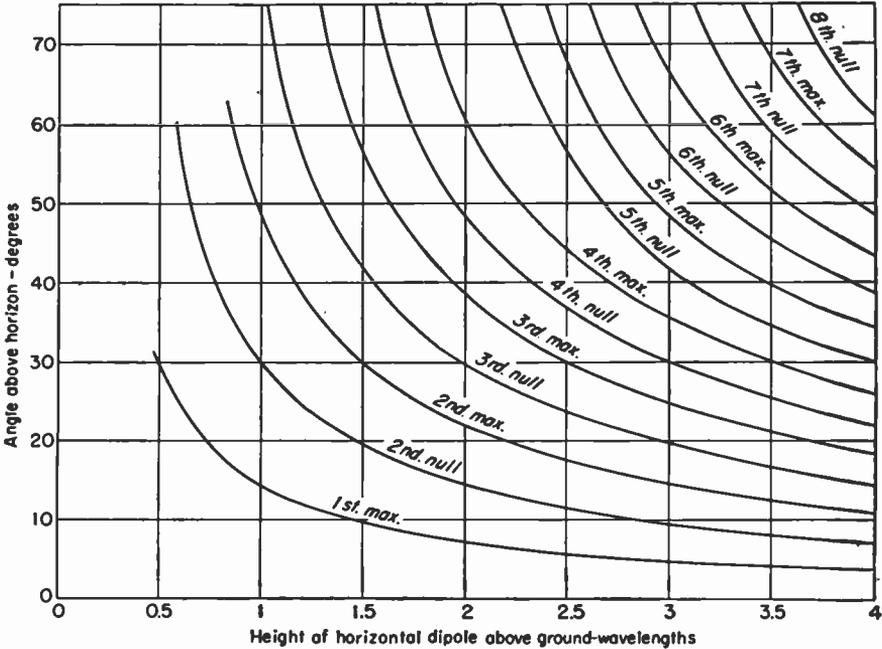


FIG. 20. Angles of maximums and minimums, horizontal dipole.

35. Radiated Field Intensities of Dipole. In free space, maximum field intensity at 1 mile ($d = 1,610$ m) with 1,000 watts radiated from a $\lambda/2$ dipole is 138 mv per m. Due to ground reflections, pattern maxima vary from one to two times this value depending upon height and ground reflectivity.

36. Selectivity and Band Width. The selectivity of a $\lambda/2$ dipole is maximum for thin wires where it can be considered to have a band width of the order of ± 1 per cent of optimum frequency. Where larger band widths are to be transmitted quite uniformly, the ratio of diameter to length of one-wire or cage antenna should be the largest practicable. If propagation considerations permit, the heights should be made such as to give maximum radiation resistances, as shown in Fig. 18. Where band widths in excess of ± 5 per cent in frequency are required, other types of construction should be used.

37. Antenna Potentials. Where high power is to be transmitted, or at high altitudes, antenna insulation and conductor designs require care to details. For h-f use, only radial potential gradients need be considered. At high altitudes, pluming may

¹ See reference 175.

occur with consequent damage to the system. Fortunately in practice, high power is generally used with directive antennas, and the power is divided among several dipole sections thus tending to minimize this problem. A thin-wire dipole gives an end potential of about 3,900 volts rms for 1,000 watts antenna input for a height of 0.25λ . It will be higher for smaller heights, and falls to a minimum of about 1,700 volts as height increases to 0.75λ ; beyond this point it settles down to the free-space value of about 3,000 volts. Potentials vary as the square root of the power ratio and as the inverse square root of the capacitance per unit length. For a potential of 3,900 volts on a wire 0.101 in. in diameter (No. 10 B&S), the radial gradient is of the

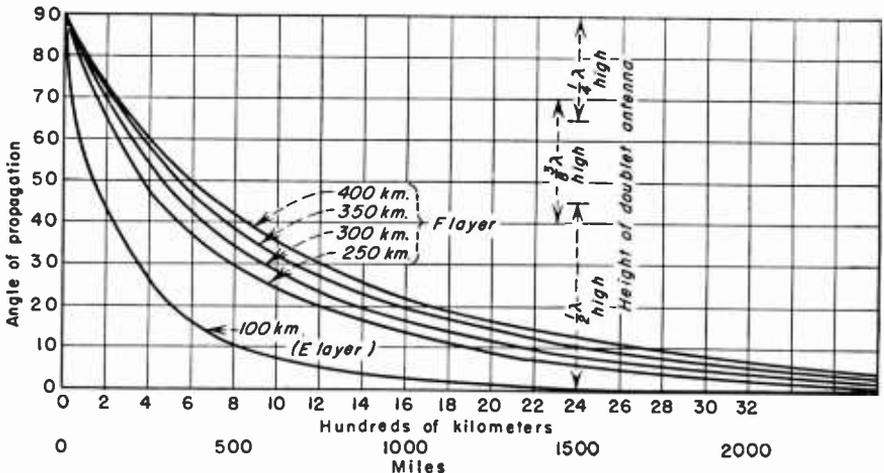


FIG. 21. Optimum radiation angle for propagation for one-hop transmission.

order of 31 kv per cm. As a rough approximation for a cage, the gradient for one wire is divided by the number of wires in the cage.

38. Half-wave Dipole for Receiving. In receiving, one wishes to know the receiver input voltage when the receiving antenna is immersed in a plane wave field of E volts per meter when both the antenna and the receiver are correctly matched to a feeder of known characteristic impedance Z_0 . This can be approximated in the following manner:

The area over which a $\lambda/2$ dipole collects energy is 0.1305 square wavelength.¹ The power within this area from the field is $W = E^2/377$. If the antenna is in free space and has no losses, the power intercepted would be, for optimum antenna orientation,

$$W = \frac{0.1305\lambda^2 E^2}{377}$$

so that the receiver input voltage would be

$$e = \sqrt{\frac{W}{Z_0}}$$

Over ordinary ground we may estimate that the effective area with the image might be about 1.9 times that in free space. Mismatches in the system, together with other

¹ See references 27 and 237.

for angle, with the pattern for the upper. In performing this addition, the polarity of the field of each lobe must be correctly included.

In free space the gain of this array is 3.95 db over a single dipole. When a plane reflecting screen is added to make it unidirectional, spaced 0.1λ parallel to the antenna, the gain can be increased by another 4 to 6 db dependent upon the area of the screen. The same principle can be used for a vertical-H array for vertically polarized transmissions.

Taking into account the actual ground reflectivity, the vertical pattern is best expressed by

$$F_{\phi} = 1 + e^{jS} \sin \phi - Ke^{i(S+2H)} \sin \phi - Ke^{i(2S+2H)} \sin \phi$$

where S = electrical spacing between upper and lower radiators

H = electrical height above ground of lower radiator

K = amplitude of reflection coefficient

Unidirectional transmission is obtained when a close-spaced reflector grid is placed on the blind side of the system.

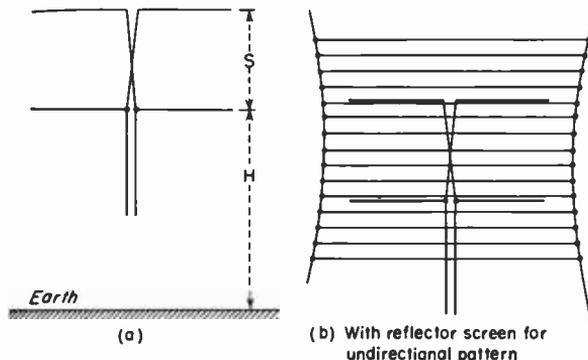


FIG. 23. Lazy-H antennas.

41. Vertical Antenna. Vertical $\lambda/2$ dipoles are not much used for h-f services. In the past, most uses of vertical dipoles have been where the uniform omnidirectional radiation property was of prevailing importance, but quadrant antennas may be used for this purpose. Vertical polarization for short and medium distances from low altitudes and over sea water is often required, and the vertical $\lambda/2$ dipole can then be employed advantageously.

Its vertical field-intensity pattern varies with its height. The pattern equation is

$$F_{\theta} = \frac{\cos(90^{\circ} \cos \theta) \cos(H \cos \theta)}{\sin \theta}$$

where θ is the angle with respect to the antenna and H is the height in electrical degrees of the current maximum, above ground. At h.f. the complex reflection coefficient from earth or water causes a mutilation of this ideal pattern. The very low-angle intensity is greatly reduced, especially along the ground. This emphasizes another important consideration with all vertical h-f and v-h-f radiators—the nature of the earth out to considerable distances from the radiating system. The performance of identical antennas in different locations may vary greatly owing to earth differences. In the immediate vicinity of the radiator, this can be beneficially controlled by the use of a radial system of elevated, surface, or lightly buried ground wires having a length of the order of $\lambda/2$. The greater the distance between the lower

end of the radiator and the earth, the smaller the importance of the wire system in reducing system losses, and the greater should be its length if used.

42. Feeding a Vertical Dipole. The most convenient method of feeding a vertical $\lambda/2$ dipole is at its lower end which is a voltage antinode. A voltage feed method is therefore necessary. If a balanced feeder is used, a line arrangement of transforming from balanced to single-end and to high standing-wave potential is required. One method is to run the open-wire balanced line past the antenna $\lambda/4$ and short-circuit it.

The antenna is then connected to one side of this line $\lambda/4$ from the short circuit. This sets up a very strong standing wave in the feeder and equalizes the unbalance introduced by feeding from one wire only. Some distance toward the transmitter the feeder is terminated with a stub line inductance or coupled matching section. In some cases the system can be fed without impedance matching if line losses can be tolerated and the input impedance is of a value that can be accommodated by the transmitter coupling. If the antenna is elevated, the feeder wires rise to the height of the lower end of the antenna and one wire connects to the antenna end and the other remains open-circuit. The unbalance caused in the feeder can be corrected at some convenient point and terminated at another point. Center feed for a vertical dipole is mechanically difficult, but when used, the feeder should be as nearly normal to the antenna as realizable to avoid large unbalance of the feeder. In general, end feed is preferred.

Vertical antennas for direction-finding stations are usually shorter than $\lambda/2$ at the highest frequency, are rigidly held in place near ground, and have extensive radial ground-wire systems for each antenna, so as to minimize wave tilt in the vertically polarized field component used for direction finding.

43. Folded Dipole Antenna.¹ This radiator is virtually a $\lambda/2$ dipole, a cage, and an impedance transformer built into one unit. H-f types made of wires are exemplified in Fig. 24. As a radiator it functions like an ordinary dipole, giving the same patterns, but with transformed radiation resistance and reactance at the feed point. Greater cross-sectional area due to folding gives it wider band-width characteristics for a given deviation in impedance, and reduces potential gradients and end potentials, for a given

¹ See reference 227.

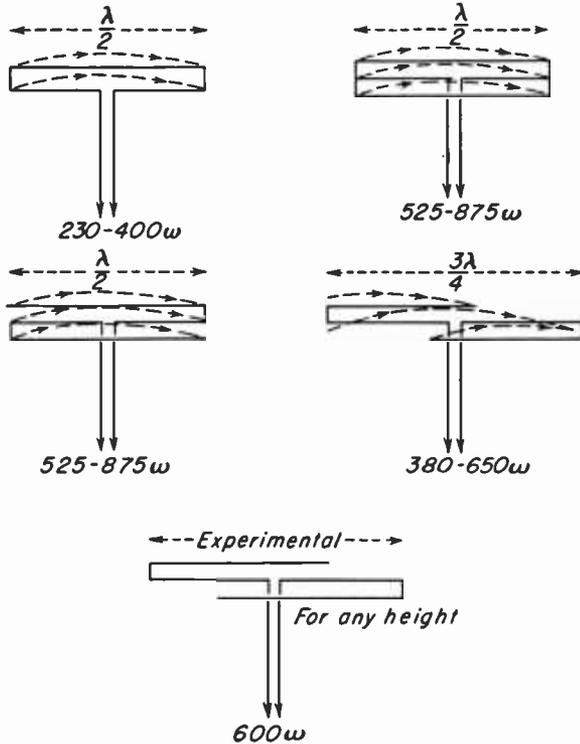


FIG. 24. Folded dipole antennas. Resistances shown are due to height above ground.

input power. Impedance transformations due to mutual impedances between the wires and due to their lengths make it possible to provide exact match with ordinary open-wire balanced feeders. A variety of combinations of wire sizes, spacings, lengths, and foldings are possible for specific applications, and single-end as well as balanced versions can be designed for h-f, v-h-f, and u-h-f uses. For a two-wire folded dipole the factor is 4, and for the three-wire system it is 9 times the radiation resistance for a single-wire dipole. The antenna impedance can be made to match the feeder by choice of spacing and wire size, height, and length.

A three-wire cage 10 in. in diameter has a band width of ± 7 per cent from optimum for a maximum of 2:1 standing-wave ratio on the feeder.

44. Quadrant Antenna (Fig. 25). The quadrant antenna¹ consists of two radiators

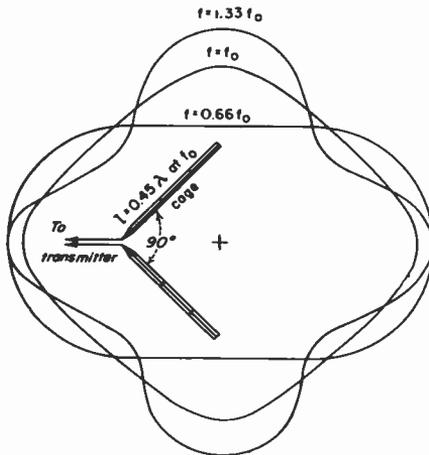


FIG. 25. Approximate horizontal field-intensity patterns for wide-band cage quadrant antenna at high-, mid-, and low-frequencies.

lying in a horizontal plane and mutually perpendicular from a common apex. Each radiator has a nominal length of 0.45λ for single-frequency use, or for the mid-band frequency when designed to obtain wide-band characteristics of approximately 2:1. Its principal feature as a radiator is its omnidirectional radiation pattern for horizontally polarized fields, as distinguished from a straight antenna. The horizontal field-intensity distribution for the latter varies approximately as the cosine of the angle to the wire, whereas with the quadrant antenna the pattern is virtually square with broadly rounded corners. The wide-band arrangement is often desirable either for eliminating antenna tuning over a 2:1 range or for the transmission of signals requiring great band width. At the extremities of the 2:1 frequency band, the pattern changes from an oval with major axis along the bisector of the quadrant and having an eccentricity of about 1.8 (lowest frequency) through the near-square pattern at mid-frequency to a quadrilobe pattern with a maximum to minimum ratio of about 1.5 at the highest frequency.

The broad-band characteristics are derived from the use of large-diameter cages having low characteristic impedance. For single-frequency operation, a single-wire design is used. The vertical pattern is determined by the height above ground just as for a straight dipole.

The quadrant antenna is fed at its apex by a balanced transmission line having characteristic impedance between 400 and 600 ohms, depending upon particular designs. Four such antennas, properly proportioned, can be used to cover the h-f band in 2:1 steps, and these can all be assembled on four supports. Identical arrangements may be used for transmitting and receiving.

45. Extensive Dipole Arrays. The H antenna is the basic form for flat arrays of dipoles which can be extended as far as desired in height and width to achieve a high degree of radiation control in vertical and horizontal planes. In the same manner, arrays can be made with dipoles in depth in any degree, in what are called *cubical arrays*. In practice, cubical arrays have found almost no use because of the complications of construction. Unidirectional patterns have been mainly obtained by placing a duplicate sheet of dipoles behind the main sheet and exciting the reflector parasitically or directly.² The adjustments of the reflector elements differ somewhat for the

¹ See reference 209.

² See references 165, 173, 174, 182, 192, and 203.

two conditions. Another simpler method is to use a close-spaced neutral reflector sheet of area somewhat larger than the radiator sheet made up of wires spaced 0.05 to 0.1λ parallel to the radiating dipoles. The closeness of spacing of the reflector wires primarily determines the magnitude of the backward radiation, and the forward radiation is but slightly increased for wire spacings smaller than 0.2λ .

Through the years a great number of distinctive beam antenna designs have appeared and disappeared. It would be possible to describe numerous ingenious forms developed in the great era of h-f growth from 1923 to 1932. In the interim, economics and the virtual abandonment of vertical polarization for long-distance h-f communications took their toll of all but a very few basic forms which we shall consider in some detail because of their importance, technically and economically. It is interesting to note that evolution has shown the absurdity of paying a high price for an extra decibel of gain when the medium of propagation varies many hundreds to one. This variation in the medium has also obscured the very poor performance of

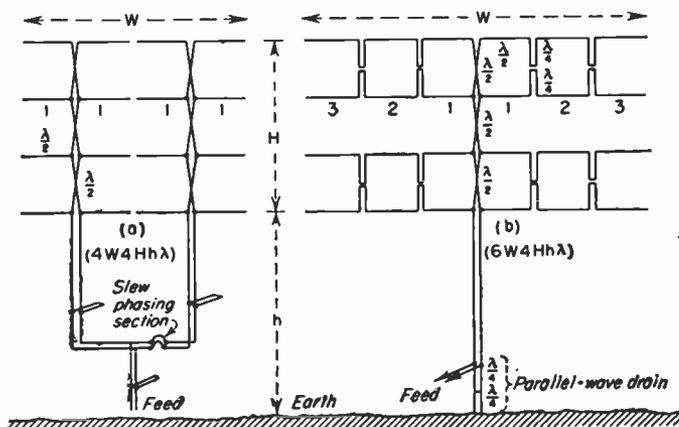


FIG. 26. Arrays of horizontal half-wave dipoles and typical feeder connections.

many antenna designs. In spite of the present stability in this field of design, work yet remains to be done to make better use of the propagation medium as more is learned of the physics of the ionosphere. Multipath transmission places a limit on the speed of keying and on speech mutilation on very long circuits. Greater control of the vertical patterns for transmitting and receiving antennas can effect improvements by angular discrimination between various orders of hops with various delays. Owing to multipath delays, the working speed decreases as circuit lengths are increased. On very long circuits, a fundamental limitation exists where the two or three lowest orders of hops come within angles which overlap those due to variation in layer height, and against which angular discrimination is impossible. In such cases the only solution seems to be to use intermediate automatic relays.

The flat array of horizontal dipoles (the pine-tree or Koomans array) is a basic type which provides independent control of vertical and horizontal patterns. The V and rhombic types have these two properties interrelated.

$4H4W$ and $4H6W$ arrays are shown schematically in Fig. 26. Dipoles are each $\lambda/2$ and spaced $\lambda/2$. Diagram A shows two feedlines, the array being two sets of four cophased double dipoles. This feeding method gives equal currents in all dipoles at one level, and there is a slight reduction in currents vertically due to attenuation along the feeders. B is a common-feeder system with three dipoles in series, cophased by the $\lambda/4$ phasing sections between colinear dipoles. Cophasing between levels is

by transposition of the feeder. Identical arrays could be used as reflectors, or a neutral sheet reflector employed.

In a multifeeder system as in *A*, the two sections of the array can be mutually dephased by a small amount to slew the beam a few degrees for different circuits with small angular path difference.

The horizontal pattern for an array of n colinear cophased dipoles arranged in m symmetrical dipole pairs from the feeder is given by

$$F_{\theta} = \frac{\cos(90^{\circ} \cos \theta)}{\sin \theta} [I_1 \cos(90^{\circ} \sin \theta) + I_2 \cos(270^{\circ} \sin \theta) + \dots + I_m \cos(90^{\circ} m \sin \theta)]$$

where the coefficient is the pattern for a single dipole, and each term in the series is the pattern for a spherical radiation source located at the center of each dipole pair symmetrically from center (feeder) and having an intensity proportional to its loop current. Azimuth angle θ is measured from the *perpendicular* to the line of the radiators. This pattern can be expressed in many other forms, but this one permits the use of current-grading coefficients as well as the usual equality of currents.

The vertical pattern for an array of n vertical elements can similarly be obtained from the following equation where each level of radiators forms a pair with its image but with image current reversed.

$$F_{\phi} = I_{11} \cos(h^{\circ} \sin \phi + 90^{\circ}) + I_{22} \cos[(h^{\circ} + 180) \sin \phi + 90] + \dots \\ + I_n \cos\{[h^{\circ} + (n-1)180] \sin \phi + 90\}$$

At an elevation angle ϕ , with a height h° between ground and the lowest radiator with current amplitude I_{11} , the first term will be the pattern for this first level. Other levels have patterns according to the successive terms at successive heights differing by $\lambda/2$. If other than $\lambda/2$ vertical spacing is used, the actual spacing is substituted for 180 deg in this equation. This equation is adapted to the use of vertical current grading by assigning proper values to I_n . Actually to grade the currents, the feeder line must act as an impedance transformer while retaining its proper phase relations at the various radiator levels.

For identical currents in all dipole elements, the maximum field-intensity gain due to directivity is proportional ideally to the square root of the number of dipoles used, both in radiator and reflector curtains. With neutral reflectors, the radiator images are counted. A number of factors cause an appreciable impairment of this ideal gain, among them ground losses, insulation losses, imperfect phase differences, reflections from discontinuities or irregularities, conductor losses, losses in associated rigging and supports, etc. A well-constructed system may be 2 to 6 db below expectation depending upon size, frequency, and location.

In constructing such a system, all conductors are cut from 5 to 10 per cent shorter than theoretical length based on a velocity of propagation equal to that in free space, to allow for effects of corner junctions, insulator capacitance, and end effect.

46. Tuning Dipole Arrays. The tuning of an array using a tuned reflecting sheet is very difficult because of mutual impedance relationships, which cause very strong interactions between radiators and reflectors. When the reflector is excited parasitically, it must be quite close to the radiator sheet to have best effectiveness. The complex nature of mutual impedances causes the spacing and adjustments to be different for maximum forward radiation as compared with minimum backward radiation. To reduce this tuning difficulty, some systems have used a radiator-reflector spacing of $5\lambda/4$ with each sheet fed directly but with a quadrature phase relationship. The objection to the latter is that additional supports are required for the reflector, whereas the close spacing permits the use of the same supports with crossarms or the use of spreaders. The simplest system is to use a neutral reflecting screen. Tuning is then confined to the one active sheet of radiators, and the spacing is diminished. If a reversible antenna is required, another sheet of radiators can be

placed on the opposite side of the reflecting screen, all using one set of supports. If both radiating systems are tuned to the same frequency, a detuning stub is switched into the idle system.

47. Dipole Array Feeders. Feeders are always kept normal to radiators to maintain correct balance, and impedance matching of the feeder is done only at one point outside the antenna system, provided all internal conditions are correct. In multiple-feeder systems, each main feeder can be terminated before being joined to the main power trunk, which in turn is matched near this junction. Care must be observed to ensure that identical conductor lengths are included in both sides of the feeder when making bends, through switches, etc. If there is radiation coupling between radiators and feeder which sets up a parallel wave to ground, the nodal points in the feeder standing-wave pattern will be displaced mutually on the two sides of the feeder. To eliminate the parallel wave, one end of a $\lambda/2$ stub can be connected across the feeder at some point and to a good ground at the other end, with a short circuit across the stub at $\lambda/4$ from the feeder junction. This stub is conveniently inserted vertically between the lower radiators and ground as shown in Fig. 26b. The main feeder can then be terminated somewhere near the end with another stub or coupled section. On a very long feeder, the cumulative effect of line irregularities caused by insulators and bends may require another matching stub nearer the transmitter.

Dipole arrays of this type, made of single wires, will transmit band widths of the order of ± 2 per cent of optimum. To broaden this band slightly, all conductors in radiators and feeders must have low characteristic impedances, most conveniently accomplished by using double-wire hairpin radiators and four-wire feeders.

An array 2W4H (eight dipoles), with lowest radiators 1λ above ground, made of 0.100-in.-diam wire, and with 8-in. porcelain insulators with $3\frac{1}{2}$ -in.-diam end fittings (insulator capacitance $1\mu\text{mf}$) has a wire length at resonance of about 0.45λ per dipole, at 18 Mc. With 100-kw carrier input, the end potential on each dipole is 7,500 volts rms. A 3,000-ft 580-ohm two-wire balanced feeder used with this antenna had a transmission efficiency of 67 per cent.

Guyed supports for h-f antennas should be removed as far as possible from the radiator fields. It must be remembered that the radiation pattern customarily referred to is that at great distance, and that very near an array there are likely to be strong fields in any direction from the radiators. The use of breakup insulators in guys is determined by their natural period. Sometimes an insulated guy will resonate with the effect of insulator capacitance between sections when an uninsulated guy would be neutral. Few or no breakup insulators are preferred when guys can be removed from strong fields and of nonresonant lengths. The current flowing in any conductor is the induced potential over the impedance. In a strong field, the impedance of a guy wire or support must somehow be made as high as possible to suppress parasitic currents.

48. Progressive Waves on Long Wires. A long wire carrying a progressive (traveling) wave has a radiation pattern following this equation in free space and without attenuation:

$$F_{\theta} = \frac{\sin \theta \left[\frac{kl}{2} (1 - \cos \theta) \right]}{1 - \cos \theta}$$

where $k = 2\pi/\lambda$

Figure 27 shows a series of polar diagrams which are cross sections of the solids of revolution surrounding wires of different lengths. The intrinsic characteristic of diminishing amplitudes for successive lobes, following the main one closest to the direction of the wire, is a valuable property in directive applications. The four sides of a rhombic antenna are each radiators of this type, with the geometry arranged to add the main lobes and further suppress the minor lobes by wave interference. The fact of attenuation present in practice has but slight influence on the major lobe, and

its principal effect is to fill in the nulls between lobes. A rigorous treatment of this type with attenuation and effects of ground has been published by Cafferata.¹

49. **Standing Waves on Long Wires.** In the same manner, a long wire in free space m half waves long with a standing-wave distribution of currents produces field distributions following these equations:²

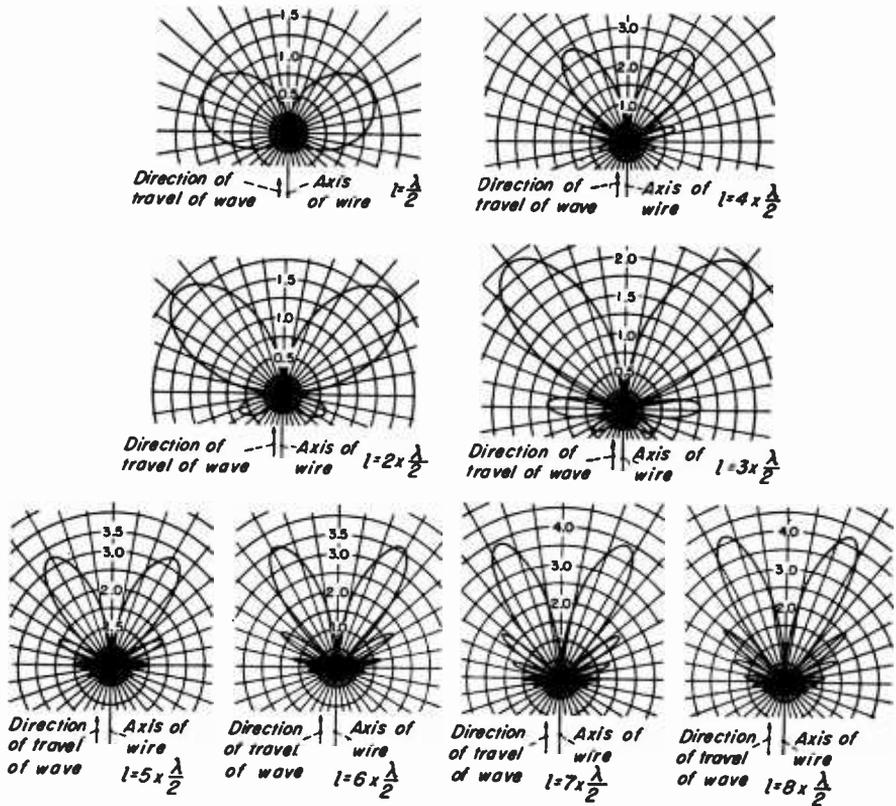


FIG. 27. Field strengths due to straight wire carrying traveling wave.

For m odd,

$$F_{\theta} = \frac{\cos\left(\frac{m\pi}{2} \cos \theta\right)}{\sin \theta}$$

Nulls occur at

$$\cos \theta = \frac{1}{m}, \frac{3}{m}, \dots, \frac{m}{m}$$

Maximums occur at

$$\frac{m\pi}{2} \tan \theta \sin \theta = \cot\left(\frac{m\pi}{2} \cos \theta\right)$$

¹ See reference 172.

² See references 1 and 173.

For m even,

$$F_{\theta} = \frac{\sin\left(\frac{m\pi}{2} \cos \theta\right)}{\sin \theta}$$

Nulls occur at

$$\cos \theta = 0, \frac{2}{m}, \frac{4}{m}, \dots, \frac{m}{m}$$

Maximums occur at

$$\frac{m\pi}{2} \tan \theta \sin \theta = \tan\left(\frac{m\pi}{2} \cos \theta\right)$$

These also neglect attenuation. The same general phenomenon as noted in the previous paragraph is present, except for the symmetry of the pattern. Figure 28 exhibits a series of patterns for medium lengths, and Fig. 29 compares directly a $7\frac{3}{4}\lambda$ and an 8λ pattern.

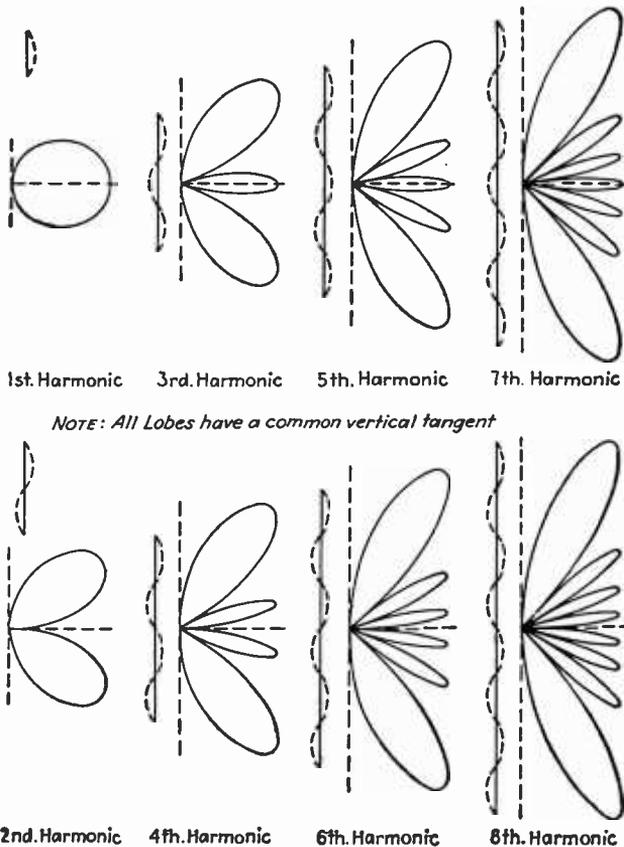


FIG. 28. Polar diagrams of relative field-strength distribution for straight wire antennas in free space with standing waves of current as shown, corresponding to the operation at various harmonics of the fundamental frequency of the antenna.

50. Trigonometric Functions of Large Angles. In computing patterns for these and other antennas where the functions of very large angles must be used, there is considerable labor in using conventional tables of the trigonometric functions because of the necessity to convert all angles to the first quadrant and get the signs correct. To aid in such work, there is reproduced in Fig. 30 a graphical chart of two-place accuracy, usually sufficient, for reading the functions directly. The versine ($1 - \cos \theta$) appears in many formulas, including those for traveling-wave patterns, and these are included in the chart along with the coversines ($1 - \sin \theta$). The tangent and cotangent functions are incomplete in such a graph.

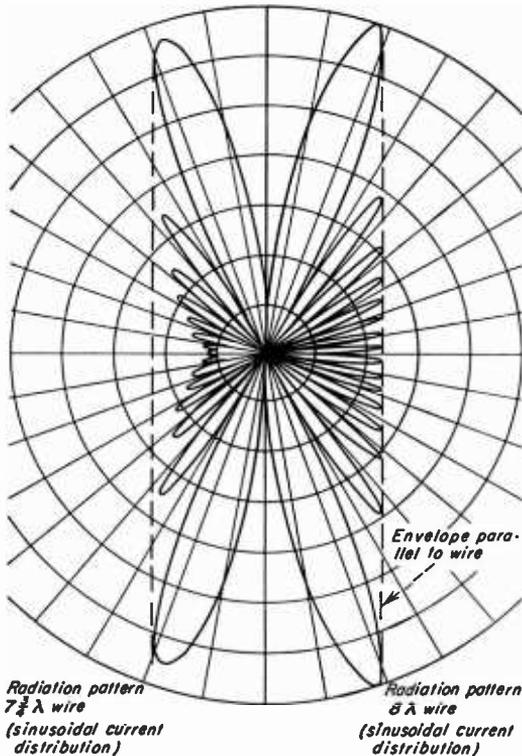


FIG. 29. Comparison of radiation patterns for standing-wave current distribution of $7\frac{3}{4}\lambda$ and 8λ wires.

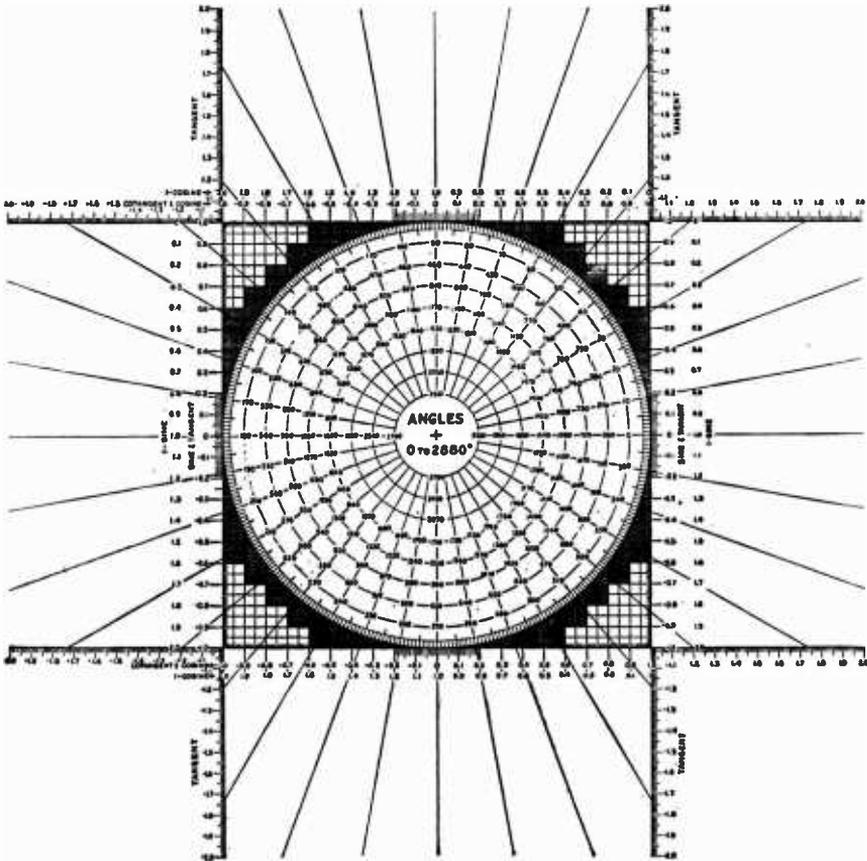
cost of r-f energy per kilowatt hour, this is not generally regarded as a serious deficiency in view of its other merits, and in relation to the fact that the propagation medium varies normally so much in proportion as to obscure this loss.

In transmitting, the potentials in the system are lower than for resonant systems. For the usual three-wire antenna with characteristic impedance of approximately 600 ohms, the maximum potential (at the input end) is approximately 7,750 volts rms for 100,000 watts carrier input. This is a desirable characteristic for high-altitude applications.

Rhombic horizontal antenna patterns are complicated by the skew geometry of the array and by the attenuated traveling waves. The main beam, usually the center of design interest, can be handled with adequate exactness by neglecting the attenuations, the complex dielectric properties of the ground, and the vertically polarized com-

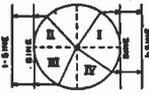
51. Horizontal Rhombic Antennas.¹ This basic type of antenna has won a permanent place in the h-f field. It is a traveling-wave system suitable for transmission and reception. It has singular merit for its broad frequency response. It is relatively inexpensive to construct and maintain and requires no tuning. Its main disadvantage is that it requires a large plot of land. A minor disadvantage is that its horizontal and vertical patterns are dependent one on the other which sometimes imposes limitations on its effectiveness. The sharper the horizontal beam, the lower the angle of the vertical beam, and it is impossible to obtain high-angle radiation except at very low gain and very broad horizontal pattern. Another disadvantage is the need to dissipate a considerable proportion of the total input power uselessly in the terminating resistance to obtain aperiodicity and unidirectionality. Except for the very high

¹ References 169, 170, 171, 188, 195, 172, 176, and 173.



DIRECTIONS:

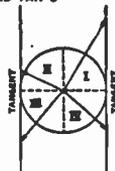
TO FIND SIN θ OR 1-SIN θ



TO FIND COS θ OR 1-COS θ



TO FIND TAN θ



TO FIND COT θ

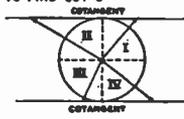


FIG. 30. Direct-reading chart of trigonometric functions of angles 0 to 2880 deg to two decimal places.

ponents of radiation. With these simplifications, the following equation¹ supplies the three-dimensional pattern:

$$F_{\theta, \Delta} = \left[\frac{\cos(\phi - \theta)}{1 - \cos \Delta \sin(\phi - \theta)} + \frac{\cos(\phi + \theta)}{1 - \cos \Delta \sin(\phi + \theta)} \right] \\ \times \sin \left\{ \frac{\pi l}{\lambda} [1 - \cos \Delta \sin(\phi + \theta)] \right\} \sin \left\{ \frac{\pi l}{\lambda} [1 - \cos \Delta \sin(\phi - \theta)] \right\} \\ \times \sin \left\{ \frac{2\pi H}{\lambda} \sin \Delta \right\}$$

where ϕ = one-half side angle of rhombus

θ = azimuth angle with respect to major axis of rhombus

Δ = elevation angle with respect to horizon

l = length of one side of rhombus

H = height of wires above ground

The axial vertical pattern is obtained by placing $\theta = 0$. The horizontal plan pattern through the maximum in the axial vertical pattern is found after substitution of Δ max in this equation and solving for variable θ .

The main beam is horizontally polarized and is directed along the major axis of the array away from the feeder end. Long arrays have a multitude of smaller beams in all directions in vertical and horizontal angles. Owing to the attenuation to the traveling waves, there are no nulls in the patterns for the individual sides of the rhombus except in the axis of the wire, which is blanked by the pattern for the other three sides. There is no vertical or horizontal angle at which there is zero radiation, even though relatively very small in magnitude.

The antenna may be regarded as a transmission line, which diverges from the feeder until it reaches an electrically large separation, then reconverges to the forward end virtually to a point. As the wires diverge, energy escapes from the system because there is no canceling effect from the opposite sides such as occurs in a close-spaced balanced transmission line. This divergence, however, is accompanied with a change in characteristic impedance unless means are taken to equalize it in relation to itself and to ground. An irregularity is introduced at the side corner so that the degree of suppression of standing waves on the terminal side is usually different from that on the feeder side. When its characteristic impedance is equalized by a spreading of three wires on each side, a more uniform match occurs throughout the system. In the case of receiving, this uniformity contributes to suppression of noise pickup from the rear. It is reported from comparative observations that a multiwire rhombic antenna is appreciably less susceptible to precipitation static interference than a single-wire system.

The three fundamental variables that control the radiation pattern are the electrical length of one side, the electrical height above ground, and the side angle at the bend. In all references, one-half of this side angle ϕ is employed as a parameter. We retain the established nomenclature for rhombic antennas herewith and assign ϕ to the half-side angle, θ to the azimuth angle with respect to the forward axis of the array, and Δ to the elevation angle. By various combinations of these factors, the main beam is directed in elevation and the beam width controlled in azimuth. At the same time the minor lobes are suppressed or enlarged in various directions so that the gain of the main beam can be slightly optimized by fine adjustment of all three parameters. Data based on an optimized design are presented in Fig. 31.

52. Rhombic Usage. The frequency range over which a rhombus can be used depends upon more than the mere fact of maintaining an impedance match which is satisfactory. Frequencies must be changed to accommodate changes in the propagation medium, and the function of a proper antenna application is to direct the radiant energy where it will be most effective. The desired radiation characteristics for various h-f circuits for various distances, times of day, season, and year, must be deter-

¹ See reference 5.

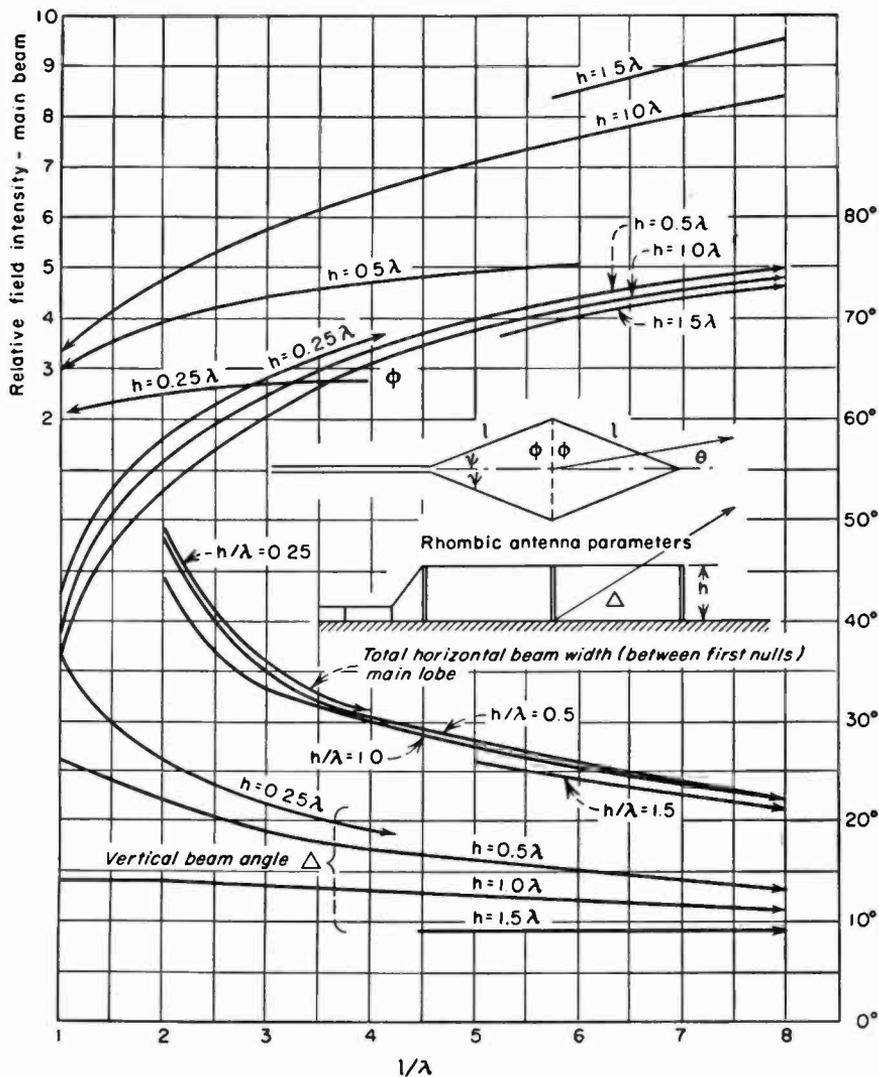


FIG. 31. Rhombic antenna design data.

mined for individual applications. Whether or not the use of a single rhombus fits the conditions encountered for the operating frequencies to be used must be investigated in relation to propagation requirements. Multipath transmission due to multiple hops and from varying heights of the different ionosphere layers imposes one of the most serious limitations on transmission speed and fidelity, and the vertical directivity of the transmitting and receiving antennas has much to do with multipath discrimination. Careful engineering of a propagation circuit may require different antennas for different frequencies regardless of the ability of a single antenna to accommodate a wide range of frequencies from an impedance-match standpoint.

In ordinary applications it has been customary to employ one rhombus for all frequencies in one direction, letting the pattern change as it will and accepting random effects in propagation. Where cost is more important than performance, this is the best that can be done. In cases where optimum use is made of the propagation medium and where performance is the primary objective, it is often best to use a different antenna specially designed for each frequency for each point-to-point circuit. When this is done, the transmitting antenna need not be aperiodic, and the energy normally thrown away in a dissipative termination can be utilized by conducting it back to the input end via transmission lines which have the correct characteristic impedance, and phased and transformed to reenter the system in parallel with that

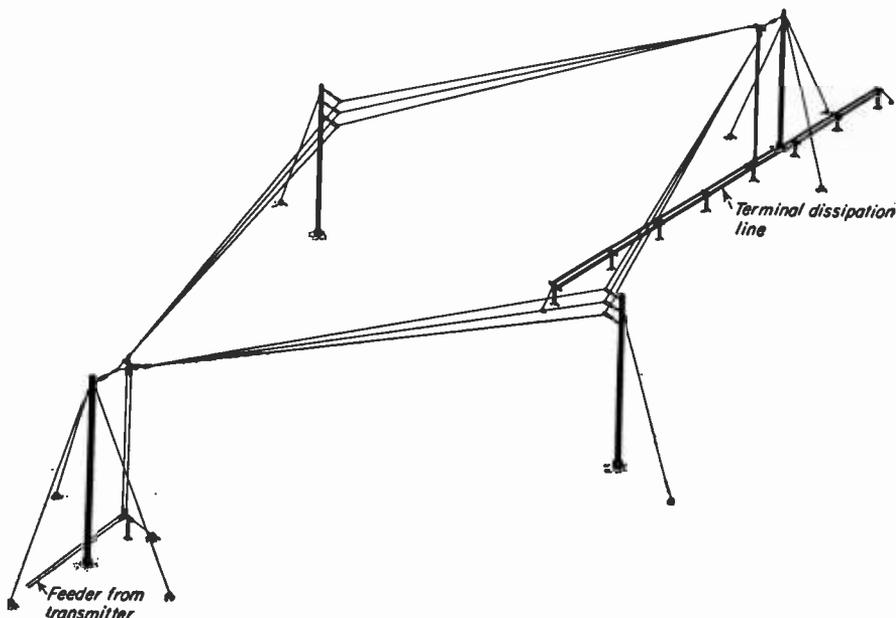


FIG. 32. Horizontal rhombic antenna.

from the transmitter. This reentrant feed method then improves the efficiency of the system for the operating frequency for which it is designed.

53. Receiving Rhombic Antennas. For receiving, either for single- or multifrequency operation, the antenna is resistance-terminated with a value as close as possible to its average characteristic impedance. Heavy-duty noninductive resistors are used, since small ones are frequently destroyed by induction from lightning. A static drain circuit to ground from the neutral of the termination is desirable to prevent the accumulation of static charges during precipitation. For transmission, where the dissipation is appreciable, high-loss transmission lines are used for terminating the antenna. Lines made of iron or stainless steel are commonly used and are assembled on low supports underneath the antenna, back and forth, to the required length. The length should be that which will provide at least 20 db attenuation at the lowest working frequency. Such a length is sufficient so that the reaction on the antenna is negligible whether open- or short-circuited. The latter is preferred because it can be directly grounded, thus serving as a static drain circuit also. The characteristic impedance is computed from formulas used for balanced two-wire lines. The attenua-

tion per unit length is given by the formula

$$\alpha = \frac{54.8 \times 10^{-6} (\sqrt{\rho\mu f/d})}{Z_0} \quad \text{db/m}$$

where ρ = resistivity, microhm-cm

μ = magnetic permeability

f = frequency, Mc

d = wire diameter, cm

Z_0 = characteristic impedance, ohm

For the line to appear as a virtually pure resistance (phase angle of impedance less than 7 deg) the $\omega L/R$ per unit length at the lowest frequency must be less than 0.1. The inductance per loop meter is

$$L = \frac{Z_0}{3} \times 10^{-8} \quad \text{henrys/m}$$

A typical layout of a three-wire rhombic antenna with dissipation and feeder lines is shown in Fig. 32.

A rhombic antenna is a balanced system, and care must be taken to preserve good balance to ground. If a number of systems are located in the same plot, sufficient spacing must be allowed to maintain low mutual impedances between the nearest sides of adjacent antennas. When two rhombics are used in broadside as a single system, one common support can be used without balance problems, when fed symmetrically. When used with concentric transmission lines, care must also be taken to make a correct balanced to single-end transformation.

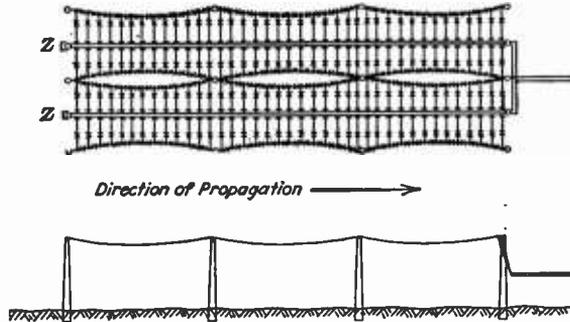


FIG. 33. Double fishbone antenna.

54. Fishbone Receiving Antennas.¹ The fishbone antenna (Fig. 33) is an evolution of the wave antenna, adapted to the reception of horizontally polarized waves in the h-f band. Three different designs are used to cover the range 3 to 20 Mc, the differences being in the lengths of the dipoles. Fishbones are used singly, or with two in broadside. A typical system is 312 ft long. A single-bay system requires 10 poles and a two-bay system 16. The essential information on this type of antenna for construction purposes is tabulated below:

Length, doublets, ft	Optimum freq., Mc	Useful range, Mc	Width, 2 bays, ft	Length, total, ft	Best pole height, ft	Useful azimuth angle, deg
34	18	13-22	120	312	60	10
48	14	10-19	148	312	90	10
66	9	3-13	200	312	120	14

The schematic arrangement of the fishbone antenna is shown in Fig. 33. A central transmission line runs horizontally in the direction of the incoming signals. At equal intervals along this line, horizontal doublets are attached which are short enough to be nonresonant within the frequency range of the system, and near enough together to

¹ See reference 166.

provide a uniform loading of the center line. Each pair of doublets loads the line with parallel capacitance and resistance. To keep the propagation velocity of the line above 90 per cent of free space, each doublet has a series capacitance near the point of attachment. The characteristic impedance of the main feeder is approximately 400 ohms. Two bays in parallel give an impedance of approximately 200 ohms, which can be matched directly into a balanced four-wire line to the receiving station. The end of the feeder toward the transmitting station is terminated in a

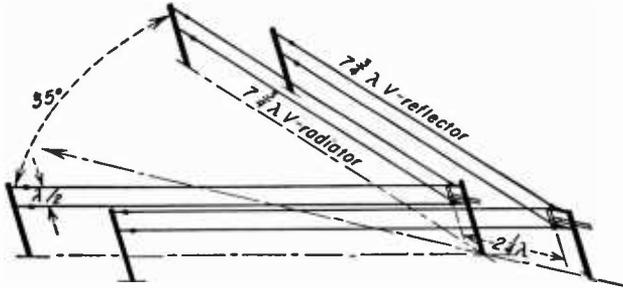


FIG. 34. V antenna with reflector.

noninductive resistance equal to the characteristic impedance of the feeder. This makes the system unidirectionally responsive by absorbing energy received from the rear. The fishbone antenna is a high-gain, broad-band, moderate-cost antenna with minimum land requirements. Even though the design is several years old, it remains one of the best receiving antennas for h-f point-to-point communication services.

55. Harmonic Wire Antennas of V Form.¹ Radiating systems using long wires having essentially sine current distributions have been called *harmonic wire antennas*. Several forms have been evolved, of which the two shown are typical (see Figs. 34 and 35). They have moderately high gain and relatively low cost. The main beam, along the axis of the V, is horizontally polarized. The vertical pattern can be con-

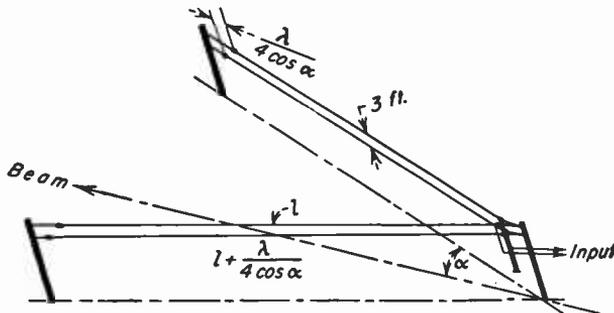


FIG. 35. Unidirectional V antenna.

trolled to a certain extent by the height of the system above ground, but this is also related to the horizontal pattern in that the sharper the horizontal pattern, the lower the angle of the vertical pattern.

Figure 29 shows the comparison of the field-intensity patterns surrounding harmonic wires having lengths of $7\frac{3}{4}\lambda$ and 8λ . The main lobe nearest the wire axis contains most of the total radiated power. V antennas are arranged to utilize this main lobe from all wires of the system so that they add along the array axis and cancel as much

¹ See reference 173.

as possible in all other directions. In Fig. 34 the two wires forming each side of each V are cophased, and the two sides of each V are antiphased. The rear V is fed 90 deg in advance of the forward V with equal currents to provide best reflecting action.

Figure 35 shows the simplest type of V antenna having unidirectional pattern. The lower wire in each side is longer than the upper by an amount equal to $\lambda/4$ in its projection upon the axis of the V, and is excited parasitically from the upper wires which are energized from the main feeder. Excited in this manner, the two wires on one side are approximately the equivalent of a single wire with a traveling-wave current distribution, producing a unidirectional pattern.

V-H-F AND U-H-F ANTENNAS

Developments in communications, f-m and television broadcasting, radar, navigational aids, and many other applications have given birth to an endless variety of antennas during recent years. This treatise cannot possibly include reference to all of them though certain basic types must be included. From these, other types can be understood. The mechanical design of antennas in these bands departs radically from those for lower frequencies because supported wires usually are replaced by rigid members, many insulators are eliminated, metallic line sections are used as insulators and in other ways. Some antennas must respond uniformly to very large band widths; some must have very high directivity; others must have switchable patterns, rotating or waving patterns, and other special properties. Since such antennas are often located in high places subject to severe weather, designs must be extremely robust. Most forms are prefabricated before erection. Electrically some are very complicated, with ingenious forms of radiators and feeding systems. Some use sheet distributions of currents instead of linear distributions, magnetic instead of electric dipoles, and many use flat, angular, or parabolic reflecting screens. In the u-h-f band, wave-guide feeders are frequently used instead of ordinary transmission lines. All of the antenna technology found in all the lower frequencies is employed in these bands together with many other principles peculiar to these bands.

56. Electric Dipoles. A description follows of certain basic radiator types used alone or in antenna arrays (see Fig. 36).

Straight Electric Dipole (Fig. 36A). A straight rod or tube of length approximately $\lambda/2$, mounted with a metallic rod attached at the mid-point (zero potential) and normal to it. To increase its band width, its diameter is made relatively great, or it may be ellipsoidal instead of cylindrical in extreme cases.¹ It can be shunt-fed with a balanced transmission line.

Figure 36B is a center-fed dipole supported by a $\lambda/4$ transformer. One side of this support acts as a continuation of the concentric feeder and, where the inner conductor emerges at the center of the radiator, it is connected to the other side. This arrangement uses the $\lambda/4$ supporting insulator also as a single-end to push-pull transformer to provide the required balanced feed potential at the center of the radiator.

Figure 36C is a form where both sides of the $\lambda/4$ support insulator act also as concentric feeders. In this case, charges flow out from the inside of the outer conductors of the feeders, propagate along the outsides of the transformer and the radiator where they set up standing waves in the usual way. The inner conductors for the two feeders are connected together.

Figure 36D shows a method of dipole construction by which the impedance at the center point of feed can be made resistive in any ordinary amount required for correct balanced feeder matching.² The $\lambda/4$ supporting transformer can be shunt-fed from a balanced line, as at points B, or the radiator can be excited directly when the feeder is attached at points A. The charges that flow on the outside of the tubular conductors set up the usual standing wave of current. The length is made somewhat less than $\lambda/4$ each way from center so that, at the feed point, the impedance appears as

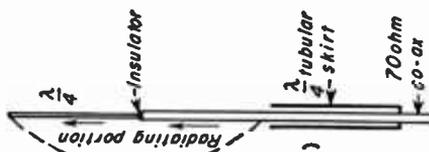
¹ See reference 249.

² See reference 223.



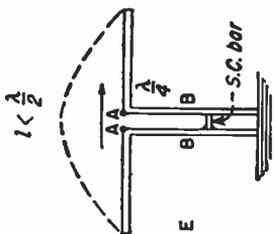
F

Co-axial antenna (cross section)

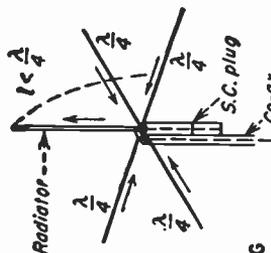


H

(Cross section)

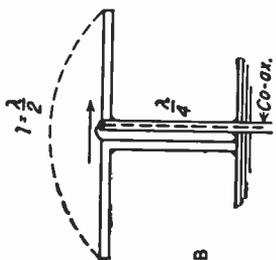


E

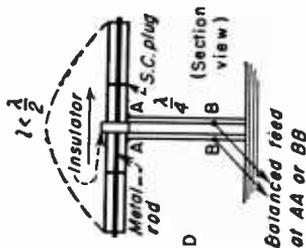


G

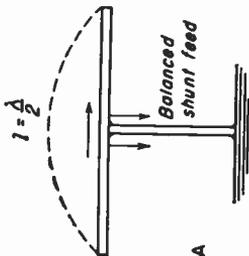
Ground screen antenna



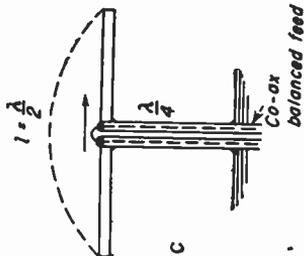
B



D



A



C

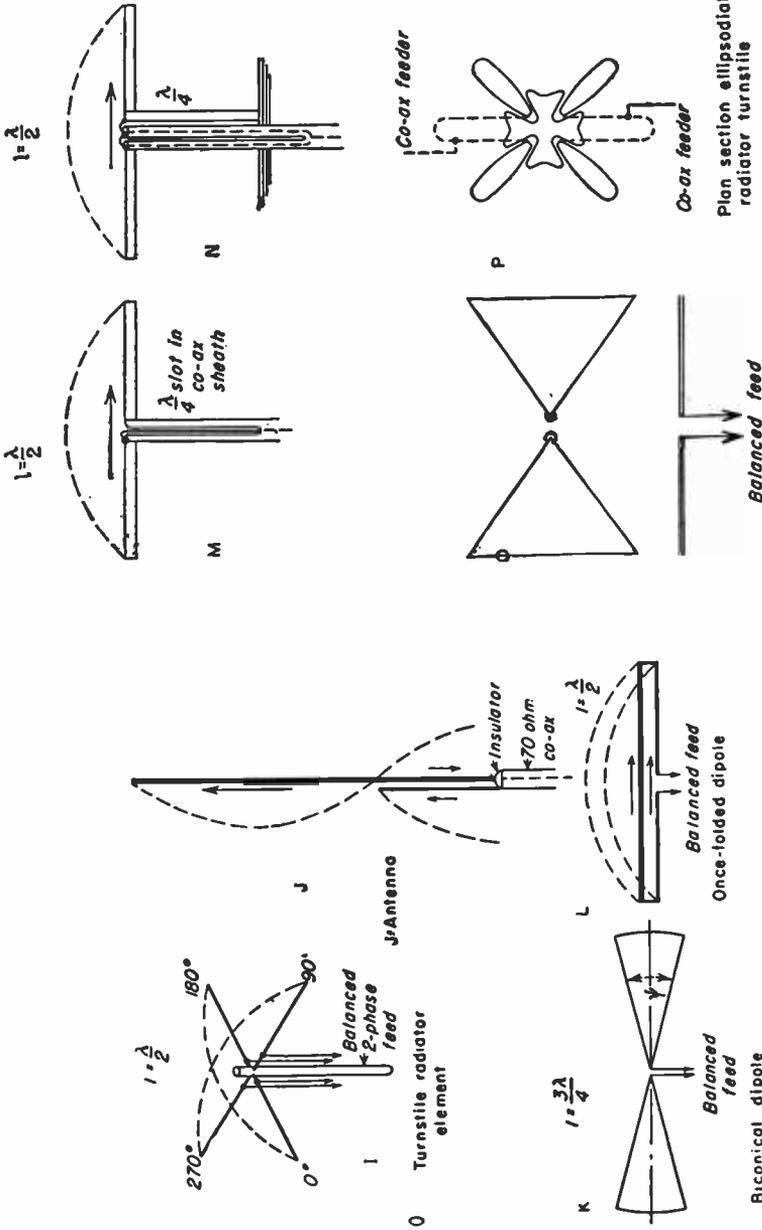


Fig. 36. Basic antenna dipoles used alone or in array.

a resistance in series with a capacitive reactance. Currents propagated inside the tubes, by virtue of the concentric conductor in the inside acting as a transmission line (and also as a mechanical support for the two halves of the radiator), are reflected from short-circuiting disks, which make the inner lines appear as inductive reactances. When these are adjusted so that this reactance forms a parallel-resonant combination with the external impedance of the radiator itself, the feed impedance is made resistive. By proportioning the external length and the internal short-circuit position, the feed-point resistance can be adjusted over a wide range to a desired value.

Figure 36E uses somewhat the same principle as the previous form, except that the inductive tuning to correct the power factor of the radiator portion is obtained by placing a short circuit across the $\lambda/4$ support transformer at the correct position. It can be fed with a balanced feeder at points *A* or *B*.

These five forms are excellent examples of multiple utilization of all parts of the system for electrical and mechanical purposes. Those using the $\lambda/4$ transformers for supports are ordinarily attached to reflecting sheets, or to some structural members for an array of such radiators. They lend themselves to virtually all-metal welded construction with a minimum of insulation.

Figure 36F is a form frequently used to obtain a vertical radiator which terminates a concentric feeder.¹ The outer conductor is, in principle, turned back $\lambda/4$ while the center conductor continues outward for $\lambda/4$. This gives a vertical dipole in which the image charges from the inside of the sheath balance those flowing upward in the projection of the inner conductor. The skirt forming the lower half of the dipole is insulated from the sheath of the feeder and suppresses sheath currents on the feeder below it. Since the feed-point impedance of such a radiator is very nearly that common to many types of coaxial line, a satisfactory impedance match and single end to balanced transformation are achieved simultaneously.

Figure 36G is a radiator less than $\lambda/4$ long that projects from a short coaxial stub-section which is short-circuited near its lower end.² The antenna presents an impedance of resistance and capacitive reactance which is tuned to a resistive condition at the feed point at the upper end of this stub, acting as a tuning inductance. Two or four horizontal $\lambda/4$ rods project outward from the lip of the concentric stub, which is attached to a mast. The impedance at the lip is adjusted by radiator length and stub length to be correct to match a concentric feeder attached to the mast. The radial rods carry the image charges for the radiator portion and by virtue of their $\lambda/4$ length, bring a potential node at the transmission line sheath. This suppresses currents on the sheath below. In this way a perfect impedance match can be made, and the assembly is compact and rugged.

The forms in Fig. 36H and *J* do the same thing in slightly different ways. In the former, the suppressor section, in the form of a cylindrical skirt connected to the outer sheath only at its lower end and being $\lambda/4$ long, presents a very high impedance at its upper edge. Thus the flow of charges from the inside of the sheath as they emerge from the feeder are reflected from this upper edge as from an open circuit. The standing waves set up on the outside of the sheath over it are images of the charges flowing in the projection of the inner conductor which is also $\lambda/4$ long. In the *J* form, the same thing takes place except that the stopper section is simply a parallel conductor forming a $\lambda/4$ transformer with the lower part of the projected center conductor which transforms the end-feed impedance back to the coaxial feeder with a value equal to that at the middle of the $\lambda/2$ radiating portion. These three forms of coaxial antennas are specially adapted to vertical antennas which must be mounted at the top of a high flagpole type of support.

Figure 36I, while composed of two straight dipoles in quadrature and excited in phase quadrature, is in reality a basic form of radiator known as the *turnstile*.³ When adjusted for equal power and exactly 90° phase difference between the two quadrature

¹ See reference 256.

² See references 222 and 235.

³ See reference 221.

dipoles, and the radiation pattern is virtually circular. As a single element, its maximum radiation is normal to the plane of the radiators. When two or more such radiators are stacked at $\lambda/2$ intervals and successively cophased, maximum radiation occurs in the plane of the radiators. When successive layers are antiphased, directivity normal to the planes of the turnstiles is obtained.¹

When this element is variously adjusted for power division and phase difference, the radiation pattern in the plane of the radiators becomes oval, the radiation is elliptically polarized, and the pattern in the plane of the turnstile is an ellipse.

A turnstile element has a certain advantage in providing a greater band width than that of the individual dipoles. This comes about from the method of feeding and phase-shifting whereby the feeder to each radiator is terminated, and the two sections excited from a common generator by means of parallel lines from a junction point with the main feeder, one branch being $\lambda/4$ longer than the other. This extra $\lambda/4$ section of line, in addition to providing the 90-deg phase lag, acts as an impedance-inverting section. Both radiators change impedance equally as the frequency is changed from mid-band, but the impedance inversion in one feeder causes a compensating effect at the main junction with the other feeder. This is a valuable property per se.

Figure 36K represents the biconical antenna whose principal special property is intrinsically greater band-width response than ordinary radiators.²

Figure 36L presents again the folded dipole. As an impedance transformer and for its greater cross section providing broader band response than a simple straight dipole, it has special value in certain v-h-f and u-h-f applications.³

Figure 36M is a variation in principle of Fig. 36B, where the central supporting member is a tube forming the outer conductor of a coaxial transmission line. The tube is slotted to a depth of $\lambda/4$, and the inner conductor of the transmission line attaches to one side of the $\lambda/2$ dipole welded to the tube.

Figure 36N is an example of a dipole, capable of many variations, where linear conductors are replaced by plane sheets of selected shapes for obtaining special characteristics.

In all these forms of electric dipoles, band width can be increased by increasing the radiator cross section when the application is associated with wide-band systems. Their radiation patterns are individually quite similar to any simple straight dipole, except for the turnstile element.

57. Magnetic Dipoles. Certain applications, such as navigational aids, require for their success the radiation of a field which is virtually pure horizontal polarization at any orientation from the antenna. Other less critical applications also use horizontal polarization, the purity of which is not essential. To obtain virtually pure horizontal polarization a magnetic dipole can be used. There are several types in existence, and others appear from time to time. What is wanted in a magnetic dipole is a loop of fairly large electrical diameter to have a high radiation resistance, with a uniformly distributed current around it, all elements of which are in time phase. It has generally been necessary to approach this by configurations made up of bent electric dipoles having standing-wave current distributions. The various forms developed exhibit ingenuity in making magnetic dipoles from electric dipoles, with varying degrees of compromise.

Some of the most practical forms are shown in Fig. 37. In this figure, A is a square loop made up of two electric dipoles, folded at their centers and near their ends.⁴ When attached together by means of insulators at the ends, the current at the corner is not zero, though it is much less in amplitude than at the other corners where the two dipoles are center-fed in balanced antiphased relationship. It is fed by a balanced feeder at the center. This has been used in aviation navigational aids. It is characterized by a nonuniform current distribution which produces an oval pattern, and by relatively low radiation resistance because of its small electrical diameter (approx-

¹ See reference 246.

² See references 227 and 74.

³ See reference 227.

⁴ See reference 212.

mately $3\lambda/16$ per side). The dipoles are made of wide bars of metal, and the feeder lines are ordinarily shielded up to the two corners where they are energized.

Figures 37*b*, *c*, and *d* are three versions of magnetic dipoles.¹ A circular or polygon loop, sectionalized and fed as shown from a coaxial feeder, permits cophasing the currents in the successive elements for relatively large electrical diameters and provides high values of radiation resistance to be realized. As the loop is increased in diameter, the periphery is sectioned into a larger number of segments, each $\lambda/2$ long, with its own radial feeder. One-half of each radiator acts as the coaxial line for the succeeding radiator.

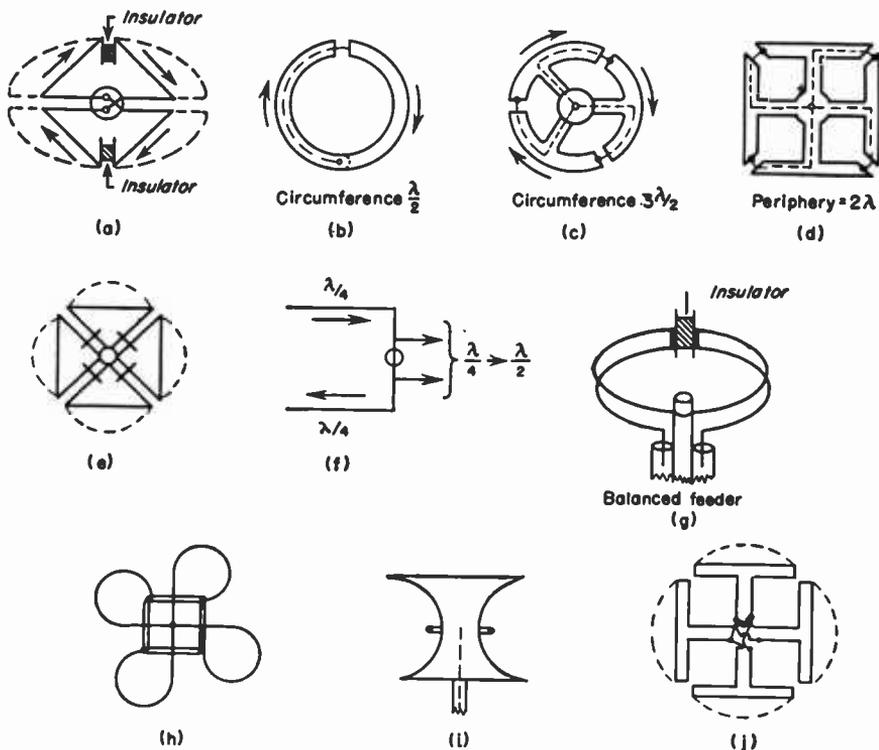


FIG. 37. Magnetic dipole antennas of various forms.

Figure 37*e* is a square (or alternately a circular) loop radiator $\lambda/2$ per side. The ends are not closed and are connected by radial supporting stubs which are short-circuited $\lambda/4$ from each corner.² Thus the current in each side is cophased, while the balanced radial stubs suppress radiation from the antiphased portions of the standing-wave system. This loop can be shunt-excited by a balanced line across one of the radial stubs, or at two such symmetrically opposite points in parallel. The pattern in the plane of the loop is squarish, the maximum field being opposite a side, and is about 1.29 times that opposite a corner.

Figure 37*f* is called the *U antenna*. A special characteristic of this radiator is that it approximates a spherical radiation source, quite unlike all other types of electric and magnetic dipoles shown.

¹ See reference 240.

² See reference 268.

Figure 37g is a folded $\lambda/2$ dipole bent in the form of a loop, using balanced feed.¹

Figure 37h is the *cloverleaf* radiator. In this case the tower itself (square) serves as the outer conductors of a concentric line, through the center of which runs the main feeder. Each cloverleaf consists of four loops of rigid tubing, each loop starting at the feeder and terminating on one corner of the tower, and bent to include a straight portion and a three-quarters circle. The length of each loop of the cloverleaf is about $\lambda/2$.

Figure 37i is a loop radiator located at the focus of a reflector which is a parabola of revolution. The reflector, depending upon its aperture, flattens the radiation pattern. The exciting loop is energized by a coaxial feeder running along the axis of the assembly.

Magnetic dipoles are used as elements in antenna arrays disposed in vertical or horizontal arrangements depending upon the application.

58. Current Sheet Radiators. Developments are progressing into the realm of exciting sheets of currents on large metallic surfaces, as distinguished from linear dis-

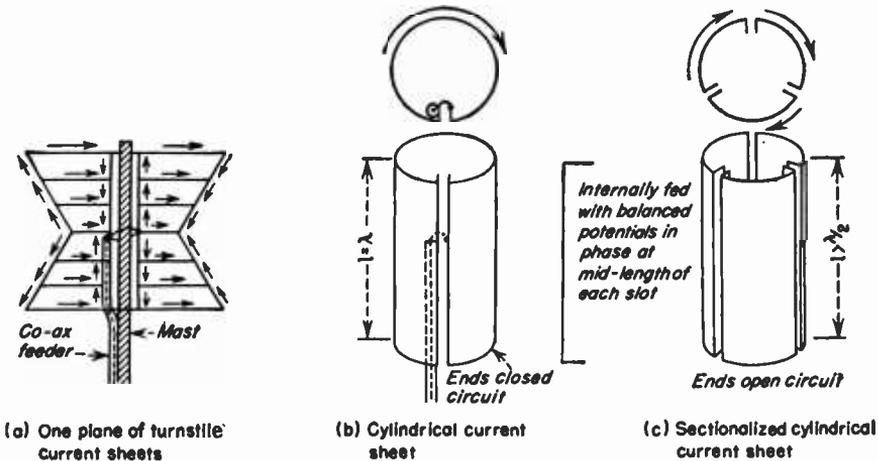


FIG. 38. Current sheet radiators.

tributions common to past practice, using wires or tubes as conductors. Current sheets may be flat or formed in various ways. Only three forms are shown in Fig. 38, as examples.

Figure 38a is a turnstile assembly of fishtail sheets in which are propagated a continuous distribution of in-phase currents flowing normal to the supporting mast. The figure shows the radiators in one plane only. This design is made of tubular members as a skeleton sheet to reduce wind surface; otherwise it functions as if made of continuous sheets. Its fishtail form was empirically developed to obtain optimum current distributions and impedance characteristics over a very broad frequency band. Each wing is attached to the mast at top and bottom, with a feeder running close to the mast but spaced from it. The horizontal elements of the skeleton sheet are welded to this and to a supporting element welded to the outer ends. A coaxial line rises to the middle of the assembly along one feeder rod to which it is connected. The center conductor emerges at this point and is connected across to the middle of the feeder of the opposite wing. This provides the balanced feed required for the radiators in one plane. The phase velocity of propagation in the vertical feeders is greater than free space velocity so that the assembly has a length of $\lambda/2$ in terms of the velocity in the system, but greater than a free-space $\lambda/2$. Two such planes of radiators in quadra-

¹ See reference 260.

ture, excited 90 deg out of phase, comprise a turnstile having a nearly circular pattern normal to the antenna axis.

Figure 38B is a current sheet bent into a cylinder having a circumference of about $\lambda/2$.¹ The slot along one side acts as the line which is excited with a standing wave by a balanced feeder, or by a coaxial feeder arranged for single-end to balanced transformation (as in the case preceding). The cylinder can be from one-half (slot ends open) to one wavelength long (ends closed) in terms of the phase velocity of a system of this type. The potentials along the feed slot are all cophased, setting up parallel current flow around the circumference of the cylinder at all points. This type of antenna has

moderate band width, very simple construction, and a pattern normal to the axis which is eccentric by ± 20 per cent. Maximum field intensity is opposite the slot.

Figure 38C is a cylindrical (or alternately polygonal) current sheet radiator having three vertical slots. Its theory is similar to the preceding example except for the three slots, the purpose of which is to obtain a pattern that is a circle. The three slots are fed in parallel at their centers by separate feeders. Measurements of the field indicate the maximum component of the electric field parallel to the slot to be less than 1 per cent of that normal to the slot. When cylindrical current sheet radiators are mounted with axis vertical, virtually pure horizontally polarized fields result. They therefore operate as magnetic dipoles, the vertical pattern being flatter, the greater the length.

59. Antennas for Radiating Simultaneous Vertically and Horizontally Polarized Fields. In certain v-h-f communication applications for aviation, it is required to transmit from the ground station signals of both horizontal and

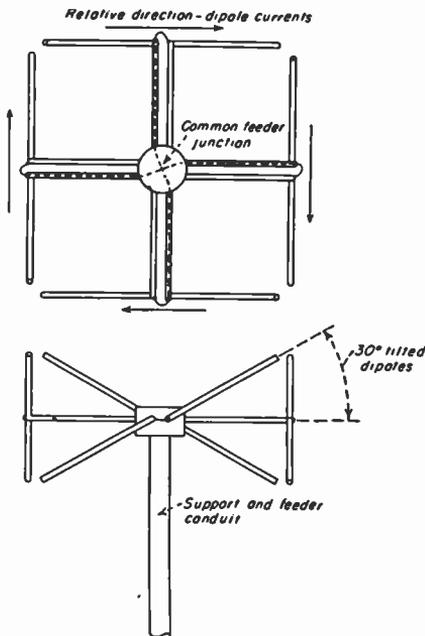


FIG. 39. Polarization antenna for v.h.f.

vertical polarization. This can be accomplished in a number of ways with separate antennas energized from a common transmitter but it is preferable to employ a single radiating system.

Figure 39 shows a type of antenna which radiates an omnidirectional circularly polarized field, especially desired in v-h-f aviation communication because of its freedom from orientation effects of the receiving antenna whether the latter be an electric or a magnetic dipole. The four electric dipoles, with equal cophased currents, form a square loop, except that the dipoles are tilted at an angle of 30 deg from their common plane, like sections of a quadruple screw thread.²

60. V-h-f Vertical Arrays. V-h-f broadcasting employs horizontal polarization (in most cases) and requires circular diffusion. To obtain the most effective concentration of radiated power toward the horizon, many forms of vertical arrays have already been evolved using magnetic dipoles, turnstile elements, or current sheets. By increasing the height of the stack, the gain can be adjusted, theoretically, to any desired value. The limitations are set by cost, feeder-line details, and the conditions encountered at the antenna location. The antenna system is always located as high

¹ See reference 287.

² See reference 290.

above surrounding terrain as possible, using a high tower, building, or natural height. In typical locations the antenna and feeder system is exposed to rigorous weather conditions, and the electrical and mechanical design must be adequate to endure them. In view of the great difficulty of making changes or adjustments after erection, such systems are now usually prefabricated and pretuned.

Gains (with respect to $\lambda/2$ dipole) for a vertical stack of magnetic dipoles fed with equal currents, cophased, and for two forms of turnstile arrays are tabulated below:

Vertical Stack of Magnetic Dipoles

Spacing, deg	Number of magnetic dipoles					
	2	4	6	8	10	12
	Vertical gain, db					
180.....	1.9	4.5	6.2	7.2	8.0	8.9
360 (optimum).....	3.2	6.3	8.4	9.6	10.8	11.7

Vertical Stack of Turnstile Elements

	Number of layers				
	2	4	6	8	10
	Vertical gain, db				
Original turnstile.....	2.0	3.5	5.0	6.5	8.0
Improved turnstile (double coaxial feed).....	1.25	2.75	4.25	5.75	7.2

Figure 40 shows the original horizontal turnstile¹ which has found wide application. Subsequent variations and improvements have been developed. Arrays of turnstiles have been employed for airways aids. One, used throughout Canada as a cone-of-silence marker on 75 Mc,² using a square array of two-layer horizontal turnstiles, excites these layers in antiphased relationship to produce a vertical beam having nearly the same beam width in all vertical planes.

Figure 41 is a form of turnstile for vertical polarization. In this form, the horizontal conductors act as balanced feeders and suppress horizontally polarized radiation. The vertical conductors are the radiators, the four in each layer, with turnstile excitation, acting to produce a nearly circular pattern in the horizontal plane.

Several commercial types of vertical arrays for f-m and television broadcasting are in existence, and others are appearing frequently, using various electric and magnetic dipoles as elements. There are arrays of cloverleaves, square loops, circular loops, circular folded dipoles, folded dipole turnstiles, etc.

Where magnetic dipoles are stacked along a common axis, the vertical pattern in free space for the array is

$$f(\theta) = \frac{\sin\left(\frac{ns^\circ}{2} \theta \sin\right) \cos \theta}{\sin\left(\frac{s^\circ}{2} \sin \theta\right)}$$

¹ See references 221 and 223.

² See reference 246.

where s° = electrical spacing between adjacent loops, deg
 θ = angle to normal from array axis
 n = total number of loops

This equation does not include the effect of ground reflections, which are usually of minor interest at receiving points along the ground. The figure of merit for an array, including the gain characteristics, depends upon the vertical pattern for the array alone. However, there will be some variation of effective power radiated in azimuth in arrays using elements having noncircular patterns.

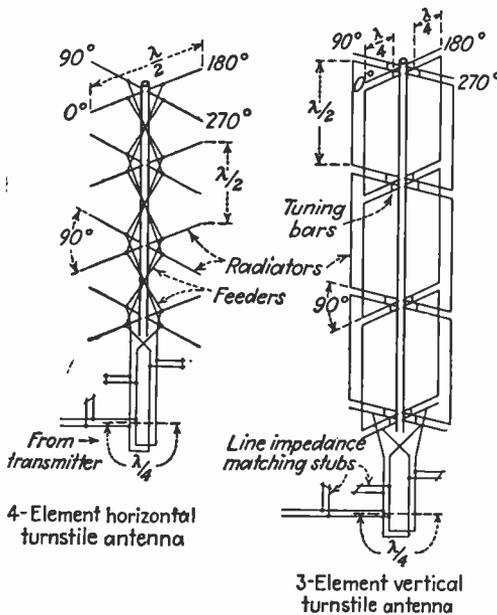
Arrays of electric dipoles may also be used for f. m., and when directive radiation is required, they will generally employ elements of this type. Current sheet radiators, singly and in vertical arrays, are also used.

The choice of f-m antennas depends upon many factors, including electrical and mechanical stability, immunity to weather, gain, band width, and, finally, effective radiated power toward the horizon per dollar of installed cost, including tower and feeders.

Television antennas are subject in general to the same requirements as f-m broadcast antennas, with the additional very important one of wide band width, with consequent standing-wave ratio on the feeder of 1.1 or less over the entire video spectrum of the channel.

When a magnetic dipole is made of horizontal electric dipoles, it is interesting to note that a triangle of three electric dipoles produces more nearly circular pattern than any other number.¹

61. Flat (Planar) Arrays. Radiation control can be effected by planar and by cubical arrays. The cubical array is unusual because it is complicated mechanically and electrically, and almost



4-Element horizontal turnstile antenna

3-Element vertical turnstile antenna

FIG. 40.

FIG. 41.

FIG. 40. Arrangement of the horizontal turnstile antenna for u.h.f.

FIG. 41. Arrangement of the vertical turnstile antenna for u.h.f.

all desired patterns can be obtained with planar arrays, often with linear arrays. However, the use of reflecting surfaces, in the form of continuous metallic sheets or metallic screens, produce in effect a shallow cubical array because of the radiation from the images in the reflector.

Flat arrays, also called *bedspring* or *billboard arrays*, have been developed in great variety for radar, most of which fall under the following types:

1. Colinear array of dipoles (Fig. 42)
2. Parallel linear array of dipoles (Fig. 43)
3. 2-dimensional (planar) array of dipoles (Fig. 44)

Depending upon the number of dipoles and their manner of excitation, the following types of free-space patterns are produced:

1. Horizontal beaver-tail pattern (Fig. 45a)

¹ See reference 228.

- 2. Vertical beaver-tail pattern (Fig. 45b)
- 3. Ellipsoidal pattern (Fig. 45c)
- 4. Lobe-switched beaver-tail or ellipsoid pattern (Fig. 46a)

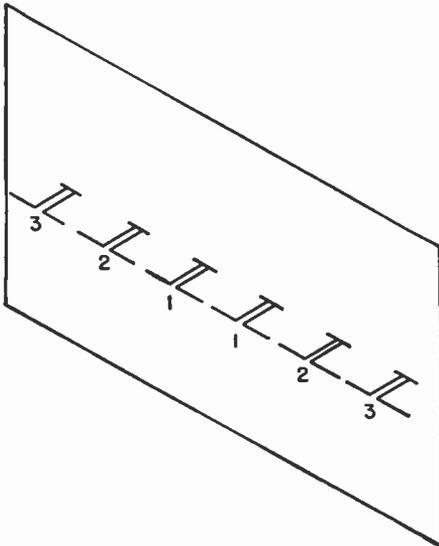


FIG. 42. Colinear array of three pairs of dipoles.

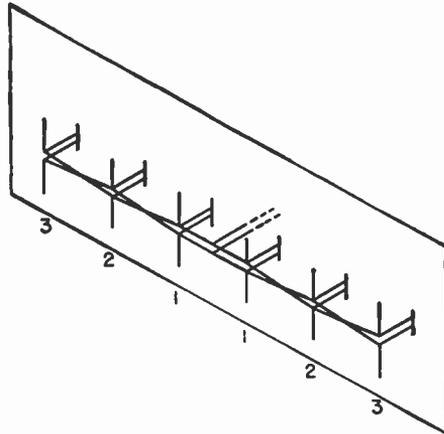


FIG. 43. Linear array of parallel dipoles.

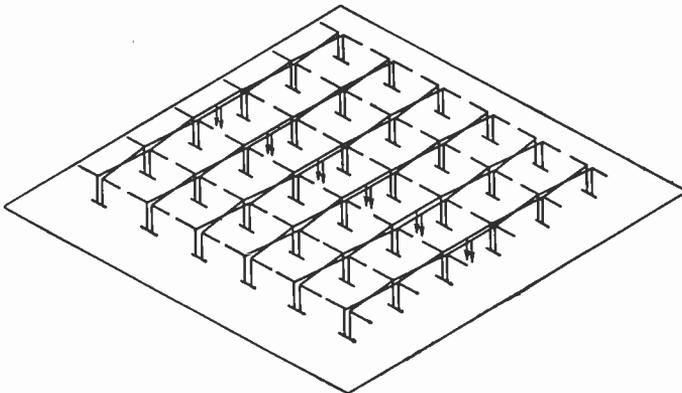


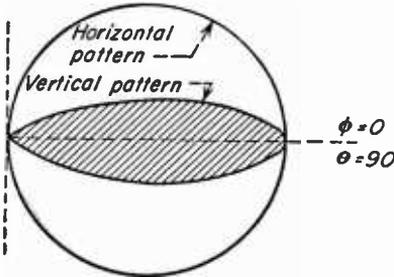
FIG. 44. Square planar array of 36 dipoles.

- 5. Ellipsoidal pattern with both horizontal and vertical switching or rotation to produce continuous conical scanning (Fig. 46b)
- 6. Forked pattern

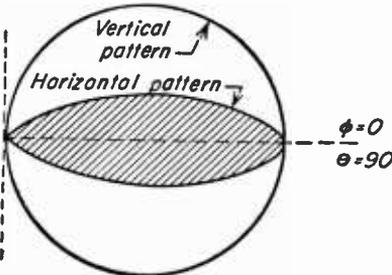
62. Pattern Shaping. Lobe switching is a process of producing overlapping patterns alternately, and using the equisignal overlap for direction finding. To obtain this, the main pattern is a few degrees from the perpendicular to the plane of the

array. The methods used to obtain a pattern of the desired shape which is slightly skewed from the perpendicular are notably the following:

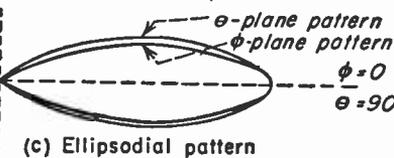
1. Dividing the array into two sets of radiators: one producing a pattern which, in terms of azimuth or elevation angle, is an even function, and the other an odd function. By inverting the polarity of the feed for the latter, the combined pattern is switched from one side to the other of the normal (Fig. 47).
2. By mechanically shifting the plane of an array periodically in one or two directions.
3. By using two separate arrays oriented to provide the desired overlap using constant excitation if fixed overlapping patterns are wanted, and switching the excitation to the two arrays if lobe switching is wanted.



(a) Horizontal beavertail pattern



(b) Vertical beavertail pattern



(c) Ellipsoidal pattern

FIG. 45. Patterns of beaver-tail antennas.

In arrays of this sort, the pairs with the greater spacings produce the greatest influence on the sharpness of the main lobe in an equal-current distribution, but they also contribute to the production of large secondary lobes. In radar applications particularly, these secondary lobes are undesirable or positively detrimental, even though the power involved may be small.

Binomial distributions are produced when the radiator currents are graded symmetrically from center as follows:⁴

- ¹ See references 26 and 79.
- ² See reference 44.
- ³ See reference 271.
- ⁴ See references 44 and 239.

Most of the arrays for communication, navigational radio systems, and radar are composed of combinations of these various characteristics. In a treatise of this scope, it is possible only to indicate briefly the principles for this vast field of design.

The patterns are shaped by the geometry of the radiating system and also by the scheme of feeding the various dipoles. The various methods of excitation for this purpose are the following:

1. Uniform currents in all radiators¹
2. Binomial current distributions symmetrically from center²
3. Fourier current distributions symmetrically from center³
4. Empirical distributions

These distributions are for the control of beam width and the elimination or reduction of secondary lobes.

The simplest arrays use $\lambda/2$ electric dipoles with half spacings of $\lambda/2$, an arrangement closely approximating a continuous sheet distribution of currents. This formula provides convenient impedances and convenient feeder systems. For the same reasons, it is desirable to treat radiators in pairs rather than singly, and the formulas given here are in that form. Other spacings can be used and occasionally are.

Array length, λ	No. pairs	Coefficients for pattern formulas					
		A	B	C	D	E	F
		Current ratios from center					
$1\frac{1}{2}$	2	0	1.00	0.33			
2	$2\frac{1}{2}$	1.00	0.67	0.16			
$2\frac{1}{2}$	3	0	1.00	0.50	0.1		
3	$3\frac{1}{2}$	1.00	0.75	0.30	0.05		
$3\frac{1}{2}$	4	0	1.00	0.60	0.20	0.1	
4	$4\frac{1}{2}$	1.00	0.80	0.40	0.11	0.014	
$4\frac{1}{2}$	5	0	1.00	0.67	0.29	0.07	0.008

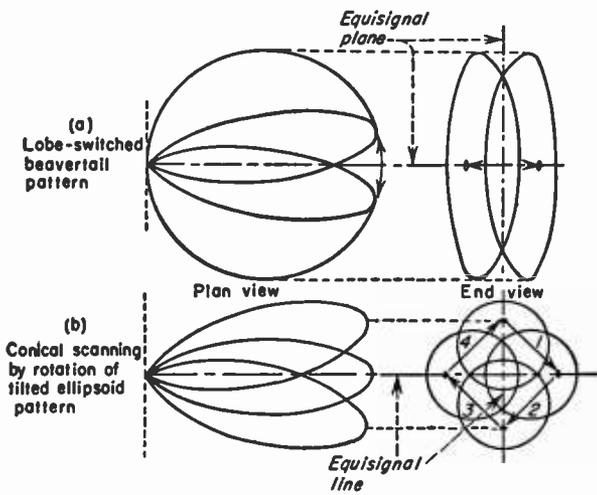


FIG. 46. Examples of lobe-switching and conical-scanning radiation patterns.

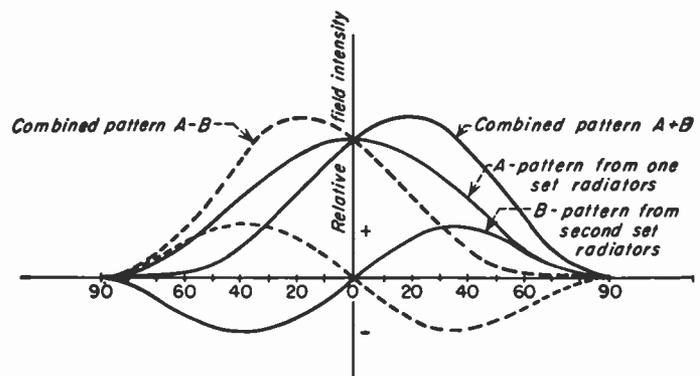


FIG. 47. Simple pattern switching by addition of odd and even pattern functions from separate sets of radiators.

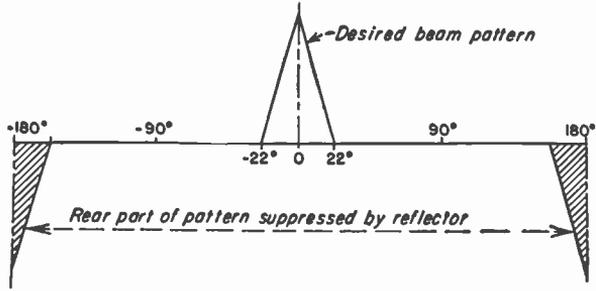


FIG. 48. Specified radiation pattern to be produced by Fourier current grading.

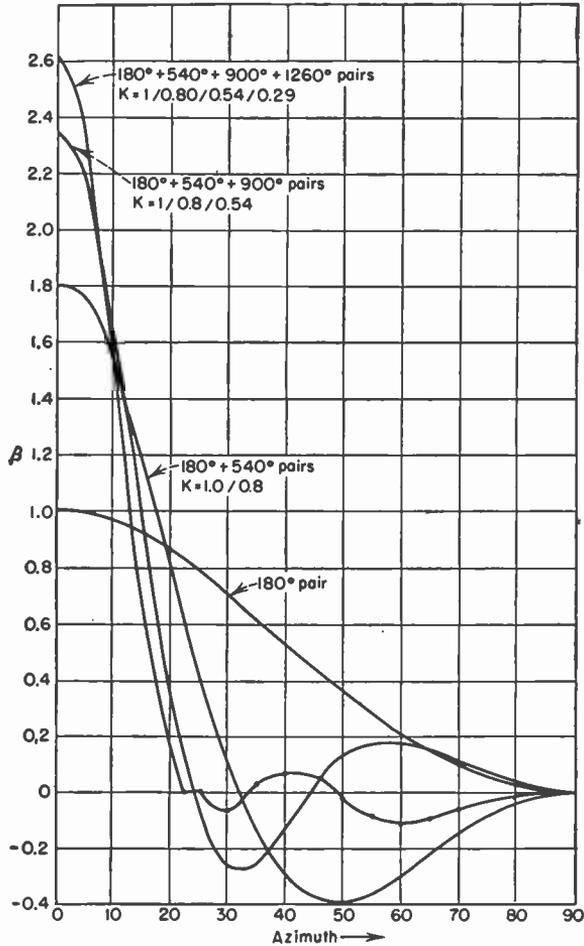


FIG. 49. Synthesis by steps of a pattern from a Fourier current distribution; four pairs.

It will be noted from this table that the current amplitudes decrease very rapidly from center. (In odd numbers of radiators, the maximum current is applied to the center radiator; in even numbers, to the central pair.) Outer pairs therefore function almost solely for side-lobe suppression and the central pairs determine the beam width. Actually, radiators with currents less than 10 per cent of maximum can be deleted with small consequences.

63. Suppression of Side Lobes. It is frequently desired to produce a particular pattern shape with side lobes suppressed to some specified maximum value where their effect is negligible. To do this, the desired pattern is Fourier-analyzed in terms of spherical harmonics, much as a distorted wave form is analyzed in terms of ordinary harmonics.¹ The coefficients from this Fourier analysis determine the current grading by pairs to be used to realize the desired pattern. The number of terms in an infinite series determines the degree of approximation that the designer wishes to approach his ideal objective.

Using spacings of $\lambda/2$ between radiators, the procedure is as follows:

Determine the current grading for a pattern which is 1.0 at 0 deg, and decreases linearly to 0 at plus and minus 22 deg, with all side lobes beyond these limits less than 4 per cent of maximum. This is a triangular pattern (in rectangular coordinates) having a base width of 44 deg. Owing to the symmetry of pattern, the calculation need be made between 0 and 22 deg only. Multiply the pattern ordinates, point by point, with the patterns for each successive pair of radiators, $\cos\left(\frac{n\pi d}{\lambda} \sin \phi\right)$ between 0 and 22 deg. Integrate each to obtain the net area after each operation. The current amplitudes in each successive pair are proportional to the areas obtained from these successive operations. This is an even function obtained by cophased pairs.

$$f(\phi) = A_1 \cos\left(\frac{\pi}{2} \sin \phi\right) + B_1 \cos\left(\frac{3\pi}{2} \sin \phi\right) + C_1 \cos\left(\frac{5\pi}{2} \sin \phi\right) + \dots$$

$$A = \int_0^{22^\circ} f(\phi) \cos\left(\frac{\pi}{2} \sin \phi\right) d\phi \quad B = \int_0^{22^\circ} f(\phi) \cos\left(\frac{3\pi}{2} \sin \phi\right) d\phi \dots$$

$$A_1 = kA = 1 \quad B_1 = \frac{B}{A} \quad C_1 = \frac{C}{A} \quad \text{etc.}$$

From this it is determined that the coefficients in the pattern equation are A, 1.00; B, 0.80; C, 0.54; D, 0.29; E, 0.09; F, 0.00, all cophased.² Figure 49 shows the result by steps of adding the patterns for the successive pairs in the amplitudes specified by the above analysis. After the addition of the fourth pair, the main lobe is a close approach to the specified shape and width, and the maximum side lobe is 4 per cent, at 60 deg. The addition of a fifth pair would reduce all side lobes to 2 per cent or less, and the main lobe would be very close to triangular. The polar pattern is shown in Fig. 50, using four pairs.

¹ See references 271 and 232.

² If there are additional terms beyond $F = 0$, there is a reversal of sign for succeeding coefficients until the next 0 is reached.

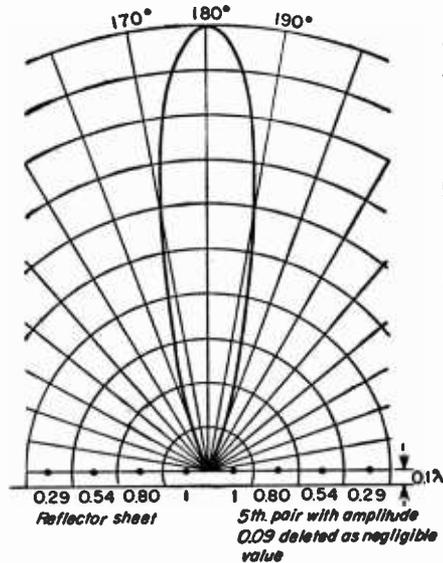


FIG. 50. Polar diagram of broadside array with current grading for side-lobe suppression and specified beam width.

Radiation Patterns for Symmetrical Arrays of Half-wave Dipoles with Half-wave Spacings, Arranged in Colinear, Linear Parallel, and Rectangular Planar Configurations, and Located a Distance X from an Infinite Perfectly Conducting Plane Reflector

1. *Equations of Pattern for m Colinear Pairs of Dipoles, Cophased* (Fig. 42):

$$f(\theta) = \left\{ \sum_1^m a_m \cos \left[\frac{(2m-1)\pi}{2} \sin \theta \right] \right\} \quad \text{Array factor}$$

$$\left\{ \left[\frac{\cos \left(\frac{\pi}{2} \sin \theta \right)}{\cos \theta} \right] \left[\cos \left(\frac{2\pi x}{\lambda} \cos \theta - \frac{\pi}{2} \right) \right] \right\} \quad \text{Dipole factor} \quad \text{Reflector factor} \quad (1a)$$

$$f(\phi) = \cos \left(\frac{2\pi x}{\lambda} \cos \phi - \frac{\pi}{2} \right) \quad (1b)$$

2. *Equations for Linear Parallel Array of n Pairs of Dipoles, Cophased* (Fig. 43):

$$f(\phi) = \left\{ \sum_1^n A_n \cos \left[\frac{(2n-1)\pi}{2} \sin \phi \right] \right\} \left\{ \cos \left[\frac{2\pi x}{\lambda} \cos \phi - \frac{\pi}{2} \right] \right\} \quad (2)$$

Array factor Reflector factor

$$f(\theta) = \left[\frac{\cos \left(\frac{\pi}{2} \sin \theta \right)}{\cos \theta} \right] \left[\cos \left(\frac{2\pi x}{\lambda} \cos \theta - \frac{\pi}{2} \right) \right] \quad (2b)$$

Dipole factor Reflector factor

3. *Planar Array of n Pairs of Rows of m Pairs of Colinear Dipoles, Cophased* (Fig. 44):

$$f(\theta) \equiv \text{Eq. (1a)}$$

$$f(\phi) \equiv \text{Eq. (2a)}$$

4. *Lobe-switched Beaver-tail Pattern for Array of Parallel Dipoles, Using n Pairs of Cophased Elements and k Pairs of Interlaced Antiphased Elements* (Fig. 46a):

$$f(\phi) = \left\{ \sum_1^n A_n \cos \left[\frac{(2n-1)\pi}{2} \sin \phi \right] \right\} \quad \text{Even-array factor}$$

$$\pm C \sum_1^k B_k \cos \left[k\pi \sin \phi - \frac{\pi}{2} \right] \left\{ \cos \left[\frac{2\pi x}{\lambda} \cos \phi - \frac{\pi}{2} \right] \right\} \quad (4)$$

Odd-array factor Reflector factor

$$f(\theta) \equiv \text{Eq. (2b)}$$

5. *Conical Scanning of Ellipsoidal Pattern for Planar Array of n Pairs of Parallel Rows of m Pairs of Colinear Dipoles, Cophased, Interlaced with k Pairs of Parallel Rows of Colinear Dipoles Having l Pairs, Antiphased* (Fig. 46b):

$$f(\phi) \equiv \text{Eq. (4)}$$

$$f(\theta) = \left\{ \sum_1^m a_m \cos \left[\frac{(2m-1)\pi}{2} \sin \theta \right] \right\} \pm D \sum_1^l b_l \cos \left[l\pi \sin \theta - \frac{\pi}{2} \right]$$

Even-array factor Odd-array factor

$$\left\{ \left[\frac{\cos \left(\frac{\pi}{2} \sin \theta \right)}{\cos \theta} \right] \left[\cos \left(\frac{2\pi x}{\lambda} \cos \theta - \frac{\pi}{2} \right) \right] \right\} \quad (5)$$

Dipole factor Reflector factor

Sequence of switching of antiphased feeders at right: *C* and *D* are usually equal. The magnitude of the odd-array factor is usually less than that of the even-array factor, by adjustment of the value of *C* or *D*.

In these equations, θ and ϕ are measured from the normal to the plane of the array. θ is the angle in the meridian plane (*E* plane) and ϕ is the equatorial plane of the dipole (*H* plane).

A_n, *a_m*, *B_k*, and *b_i* are Fourier, binomial, uniform, or other systematic coefficients for the current distributions in successive pairs of radiators to obtain desired pattern shapes.

64. Flat Reflecting Screens. A conducting screen, mesh or sheet, of sufficient area, when placed behind a radiating system, acts as a reflecting surface.¹ Use is made of this fact in v-h-f flat arrays. There is diffraction around its edges which sets up fields in the rear of the reflector, and the size used is chosen to keep the stray fields below certain values, depending upon the application. The size is empirical, but in

Interval	<i>C</i>	<i>D</i>
1	+	+
2	-	+
3	-	-
4	+	-
1	+	+

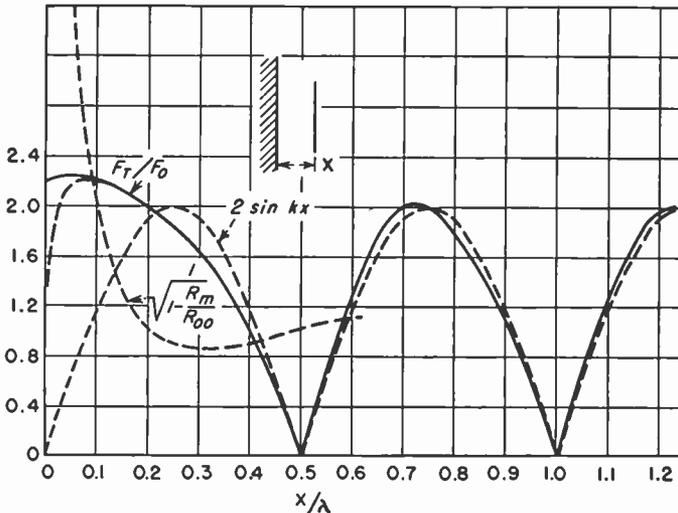


FIG. 51. Transmitting antenna in front of flat sheet. *F_r* is the field intensity with reflector; *F₀* is the field intensity from a free-space dipole; *R_m* is the mutual impedance between the dipole and its reflector image; *R₀₀* is the self-resistance of the free-space dipole.

general the higher the directivity of the radiating system, the smaller can be the projecting dimensions of the reflector. In electrically large arrays, the dimensions of the reflector have been as little as $\lambda/2$ more than the radiating system in both length and width. The reflector size depends also on the spacing from the radiators, but this is unimportant for the usual spacings of from 0.1 to 0.25 λ . When screens made of wire grids are used, the wires are parallel to the radiators. The spacing between grid wires can be as large as 0.2 λ for acceptable forward effectiveness; for better suppression to the rear, smaller spacings are necessary. Designs requiring a high degree of suppression of backward radiation use spacings as small as 0.05 λ . The same is true of square mesh screens. Screens are preferred to continuous metallic sheets for arrays exposed to winds.

Figure 51 shows the ratio of forward field with a continuous metallic reflector to that without a reflector as a function of spacing for the transmitting case. Figure 52

¹ See reference 102.

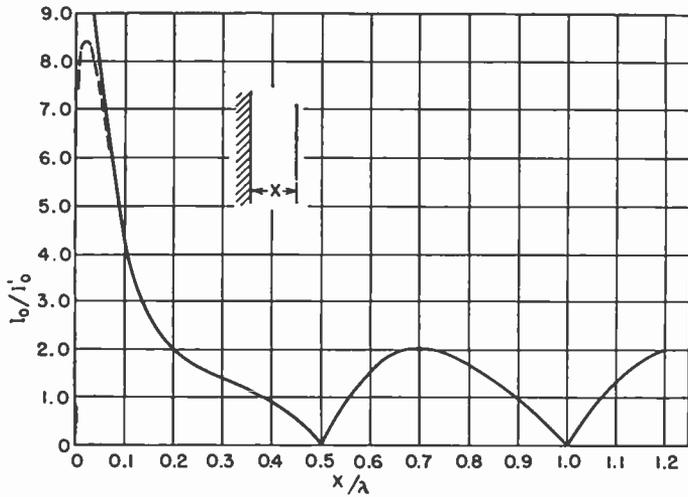


FIG. 52. Receiving antenna in front of flat sheet. I_0' = dipole current without the reflector. I_0 = dipole current with the reflector present.

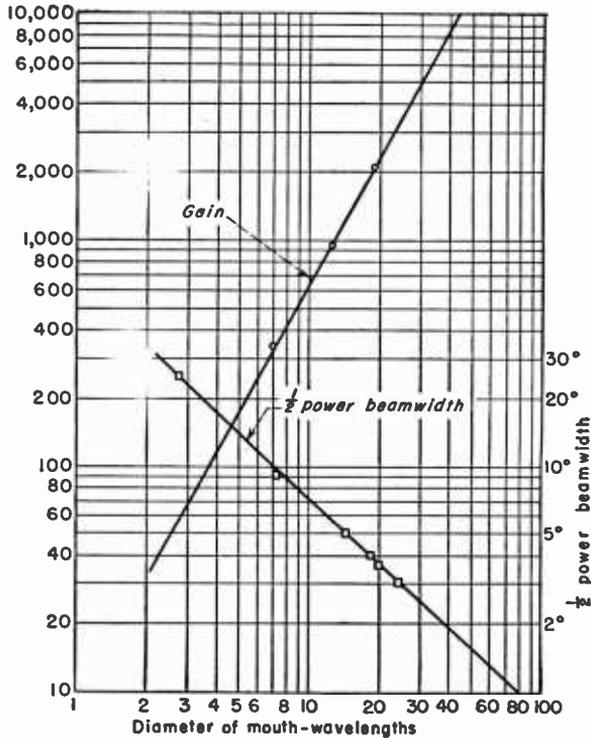


FIG. 53. Paraboloidal projector; typical measured characteristic. Power gain over $\lambda/2$ dipole.

shows the ratio of received antenna currents with and without a reflector. These approximate the results with screens and grids of finite dimensions.¹

65. V-h-f Antenna Measurements. Performance measurements on v-h-f antennas consist principally of (1) measurement of the radiation pattern and (2) measurement of the impedance at the main feeder by means of observing the standing waves on the feeder to be used with the system. For wide-band systems, these are repeated throughout the range of frequencies to be accommodated.

Pattern measurements are made by energizing the radiating system and rotating it on a turntable while observing the field intensity on a recorder at a fixed point. The field strength vs. angle is thus examined through 360 deg. Patterns at any orientation can be measured in this way. Errors of large magnitude result when the receiving equipment is moved around a fixed transmitting antenna. The location used should be open and free of objects which will produce reflections except in the direction of the receiver, and the receiving antenna should be highly directive toward the turntable and discriminate as much as possible against pickup from the sides. The amplitude response of the detector used must be known so that corrections can be computed for its curvature influence on the observed values. A linear detector indicates relative field intensities directly, and a square-law detector indicates relative power directly. Most detectors fall between these limits, and their empirical results must be corrected.

Incorrect patterns indicate incorrect amplitudes and phases of the currents in the radiators. Improper impedance matches in the feeders must then be located by analysis and corrected. Very little can be done by direct measurement in these feeders, and the final results must be achieved by measurement of the pattern.

The input impedance to the array can be measured by the slotted-line technique or other suitable impedance-measuring instrument.

The band width of an antenna from an input-impedance standpoint is normally determined in terms of some specified standing wave ratio in the main feeder, assuming virtually perfect match at mid-band. The maximum permitted at the edges of the band depends upon the intended service. For some noncritical applications this can be a standing-wave ratio of 2, while for others it must be as low as 1.1. The standing-wave ratio should be specified when specifying band width.

66. Radiators with Parabolic Reflectors. Parabolic shapes are used in various ways to produce, by reflector action, many types of patterns. Cylindrical parabolas, paraboloids, and sections of each have been widely used, especially in radar applications.

To be adequately effective, the reflectors must be quite large, of the order of wavelengths. Figure 53 shows typical values of half-power beam width and power gain over a $\lambda/2$ dipole for paraboloidal reflectors, which represent the usual design values used where side-lobe radiation is minimized.

The radiator used is ordinarily a $\lambda/2$ dipole with a disk, cup, or reflector dipole in front so as to throw most of the radiant energy into the reflector. A common arrangement for use with a paraboloidal reflector is shown in Fig. 54. By aiming the array in the manner of a searchlight, the beam is directed as desired for fixed or scanning uses.

To reduce wind resistance, skeleton forms are employed either as mesh screens or as slotted screens, the slots being parallel to the exciting dipole.

¹ Brown, G. H., *Directive Antennas*, *Proc. IRE*, January, 1937, p. 122.

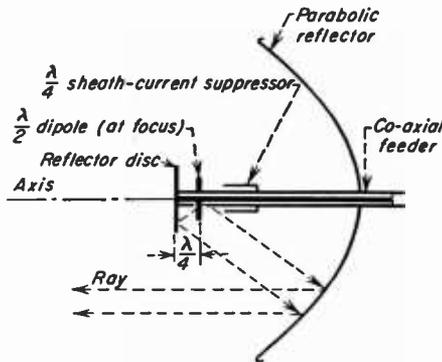
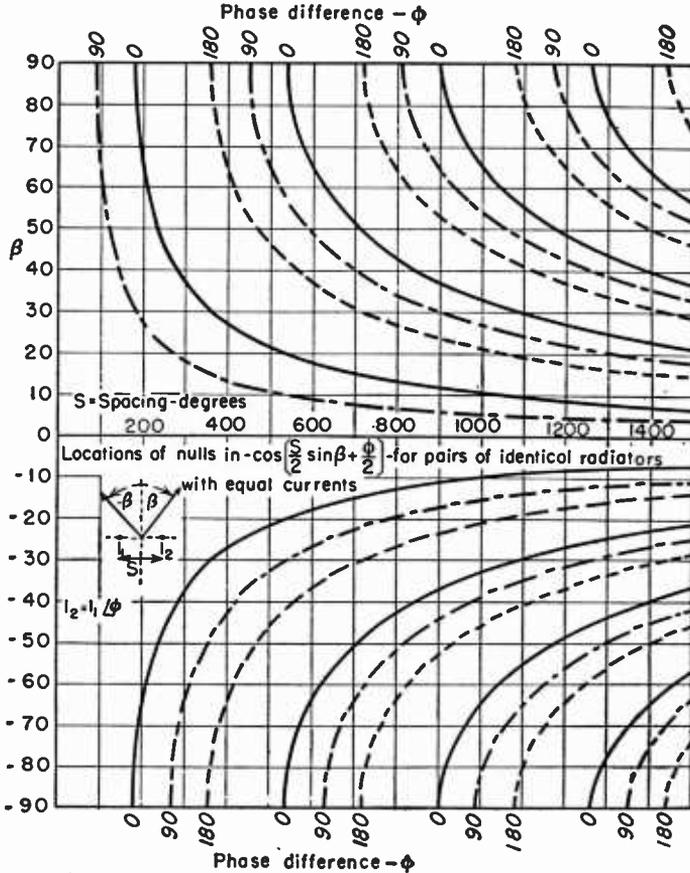


FIG. 54. Common form of microwave parabolic radiating system.

The equation for the surface of such a reflector is

$$y^2 = 4px$$

where y is the ordinate parallel to the mouth, p the focal length, and x the abscissa measured along the axis. The radiator is located at the focal point.



Note - All patterns symmetrical about line thru radiators

Fig. 55. Curves useful for determining nulls in patterns.

Where a slice (beaver-tail) pattern is desired, a parabolic section is used which may be a portion of a parabolic cylinder or paraboloid. The wide part of the section is in the plane of maximum sharpness of pattern.

Lobe switching and scanning can be done by exciting the reflector off focus, and by changing the position of the dipole as required for the desired scanning pattern.

67. Curves Showing Locations of Nulls in Patterns Such as Fig. 15. In designing directive antennas of various types, use is often made of polar patterns for pairs of radiators, and more extensive arrays are synthesized from such pairs. Figure 55 will be useful in such exploratory work because it shows the angles at which the nulls occur. Interpolation can be made easily between the three phase lines permitting

Table of Values of Functions Corresponding to Radiation Patterns for the Half-wave Dipole, Several Radiator Pairs, and Groups of Pairs. (Continued)

$f(\phi)$ ϕ	Cophased pairs of circular radiation sources							Antiphased pairs of circular radiation sources			$\cos \lambda \sin \phi + 90^\circ$ for ϕ angles all signs are reverse of those shown		
	$\cos \left(\frac{1}{2} \sin \phi \right)$	$\cos \left(\text{II} \sin \phi \right)$	$\cos \left(\frac{3}{2} \sin \phi \right)$	$\cos \left(\text{III} \sin \phi \right)$	$\cos \left(\frac{5}{2} \sin \phi \right)$	$\cos \left(\text{III} \sin \phi \right)$	$\cos \left(\frac{7}{2} \sin \phi \right)$	$\cos \left(\text{III} \sin \phi \right)$	$\cos \left(\frac{9}{2} \sin \phi \right)$	$\cos \left(\text{III} \sin \phi + 90^\circ \right)$		$\cos \left(\frac{7}{2} \sin \phi + 90^\circ \right)$	$\cos \left(\text{III} \sin \phi + 90^\circ \right)$
52.5°	0.523	0.32	-0.79	-0.83	0.25	0.99	0.38	-0.75	-0.87	-0.83	0.96	-0.83	0.51
55°	0.488	0.28	-0.84	-0.76	0.41	0.99	0.14	-0.91	-0.66	-0.59	0.90	-0.99	0.77
57.5°	0.453	0.24	-0.88	-0.87	0.56	0.94	-0.07	-0.98	-0.38	-0.46	0.83	-0.99	0.91
60°	0.414	0.21	-0.91	-0.89	0.67	0.87	-0.29	-1.00	-0.11	-0.40	0.75	-0.95	1.00
62.5°	0.380	0.18	-0.93	-0.81	0.76	0.78	-0.46	-0.86	0.13	-0.34	0.67	-0.88	0.99
65°	0.345	0.15	-0.95	-0.43	0.83	0.68	-0.63	-0.86	0.37	-0.29	0.574	-0.76	0.94
67.5°	0.309	0.12	-0.97	-0.35	0.88	0.56	-0.75	-0.74	0.57	-0.24	0.47	-0.67	0.84
70°	0.271	0.10	-0.98	-0.28	0.92	0.45	-0.84	-0.62	0.73	-0.19	0.38	-0.55	0.69
72.5°	0.240	0.07	-0.98	-0.22	0.96	0.36	-0.90	-0.49	0.85	-0.05	0.29	-0.44	0.53
75°	0.204	0.06	-0.99	-0.16	0.98	0.28	-0.94	-0.37	0.91	-0.05	0.22	-0.34	0.39
77.5°	0.171	0.04	-0.99	-0.10	0.99	0.20	-0.97	-0.25	0.96	0.06	0.16	-0.23	0.28
80°	0.138	0.03	-1.00	-0.05	1.00	0.14	-0.99	-0.15	0.98	0.04	0.10	-0.14	0.17
82.5°	0.103	0.02	-1.00	-0.03	1.00	0.09	-1.00	-0.09	0.99	0.02	0.06	-0.08	0.11
85°	0.070	0.01	-1.00	-0.01	1.00	0.05	-1.00	-0.04	1.00	0	0.03	-0.04	0.07
87.5°	0.036	0	-1.00	0	1.00	0.02	-1.00	-0.01	1.00	0	0.01	-0.01	0.03
90°	0	0	-1.00	0	1.00	0	-1.00	0	1.00	0	0	0	0

close estimation of the null positions for any value of phase difference between radiator currents.

VALUES FOR CERTAIN BASIC PATTERNS FOR PAIRS OF RADIATORS AND FOR THE PATTERN OF THE $\lambda/2$ DIPOLE

Various patterns for extended arrays are synthesized by addition and multiplication of functions for basic elements, among them the $\lambda/2$ dipole pattern, and the patterns for certain pairs of radiators. Since $\lambda/2$ spacing between elements of a pair is most frequent in practice, owing to ease of feeding an array, there are tabulated below such combinations as are usually required for cophased and antiphased pairs with integral multiples of the $\lambda/2$ spacing, and for the action of neutral reflectors. With these tables, many of the broadside beam antenna patterns can be quickly calculated, including those using various systems of current grading between successive pairs, for both odd and even functions.

References

General Textbooks:

1. Stratton, J. A.: "Electromagnetic Theory," McGraw-Hill, 1941.
2. Schelkunoff, S. A.: "Electromagnetic Waves," Van Nostrand, 1943.
3. Sarbacher and Edson: "Hyper and Ultra High Frequency Engineering," Wiley, 1943.
4. Atwood, S. S.: "Electric and Magnetic Fields," Wiley, 1941.
5. Terman, F. E.: "Radio Engineers' Handbook," McGraw-Hill, 1943.
6. "ARRL Antenna Handbook," American Radio Relay League.

Radiation and Antenna Theory:

7. Ballantine, Stuart: On the Radiation Resistance of a Simple Vertical Antenna at Wave Lengths below the Fundamental, *Proc. IRE*, **12**, 823, December, 1924.
8. ———: On the Optimum Transmitting Wave Length for a Vertical Antenna over Perfect Earth, *Proc. IRE*, **12**, 833, December, 1924.
9. Barrow, W. L.: Impedance of a Vertical Half-wave Antenna above an Earth of Finite Conductivity, *Proc. IRE*, **23**, 150, February, 1935.
10. Bashenoff, V. I., and Mjasoedoff, N. A.: Effective Height of Closed Aerials, *Proc. IRE*, **19**, 984, June, 1931.
11. Bechmann, R.: Calculation of Electric and Magnetic Field Strengths of Any Oscillating Straight Conductors, *Proc. IRE*, **19**, 461, March, 1931.
12. ———: Calculation of Radiation Resistance of Antennas and Antenna Combinations, *Proc. IRE*, **19**, 471, August, 1931.
13. Born, H.: Energy Distribution in the Near Field of Electro-magnetic Radiators, in Particular in Front of Diaphragms and Reflectors (treated by acoustical methods); abstract, *Wireless Engr.*, **21**, 188, April, 1944.
14. Bouwkamp, C. J.: Hallens Theory for a Straight, Perfectly Conducting Wire, Used as a Transmitting or Receiving Aerial, *Physica*, **9**, 609-631, July, 1942.
15. Brown, G. H., and R. King: High-frequency Models in Antenna Investigations, *Proc. IRE*, **22**, 457, April, 1934.
16. Brown, G. H.: Phase and Magnitude of Earth Currents near Radio Transmitting Antennas, *Proc. IRE*, **23**, 168, February, 1935.
17. ——— and O. M. Woodward: Experimentally Determined Impedance Characteristics of Cylindrical Antennas, *Proc. IRE*, April, 1945, pp. 257-262.
18. Burgess, R. E.: Aerial Characteristics; Relation between Transmission and Reception, *Wireless Engr.*, **21**, 154, April, 1944.
19. Carter, P. S.: Circuit Relations in Radiating Systems and Applications to Antenna Problems, *Proc. IRE*, **20**, 1004, June, 1932.
20. Colebrook, F. M.: Experimental and Analytical Investigation of Earthed Receiving Aerials, *J. IEE*, **71**, 235, July, 1932.
21. ———: Electric and Magnetic Fields for a Linear Radiator Carrying a Progressive Wave, *J. IEE*, February, 1940, p. 169.
22. ——— and A. C. Gordon-Smith: Method of Calibrating Field Strength Measuring Set, *J. IEE*, March, 1941, p. 15.
23. Cronshey, R. W.: Calculating Antenna Radiation Patterns, *Elec. Eng.*, **63**, 331, September, 1944.
24. Cutting, Fulton: Simple Method of Calculating Radiation Resistance, *Proc. IRE*, **10**, 129, April, 1922.
25. Forbes, Henry C.: Re-radiation from Tuned Antenna Systems, *Proc. IRE*, **13**, 363, June, 1925.
26. Foster, Ronald M.: Directive Diagrams of Antenna Arrays, *Bell System Tech. J.*, **5**, 292, January, 1926.
27. Friis, H. T.: A Note on a Simple Transmission Formula, *Proc. IRE*, **34**, 254, May, 1946. (Effective areas of transmitting and receiving antennas.)

28. Giaeletto, L. J.: Antenna Radiation Chart, *Electronics*, **12**, 35, July, 1939.
29. Gray, M. C.: Modification of Hallen's Solution of the Antenna Problem, *J. Applied Phys.*, **15**, 61, January, 1944.
30. Green, E.: Extended Aerial Systems; Calculating the Polar Diagrams, *Wireless Eng.*, **19**, 195, May, 1942.
31. Grosekopf, J.: Radiation Field of a Vertical Transmitting Dipole over Stratified Ground, *Wireless Eng.*, **20**, 245, May, 1943.
32. Grover, F. W.: Methods, Formulas, and Tables for the Calculation of Antenna Capacity, *Nat. Bur. Standards, Sci. Papers* 568-569, 1928.
33. Hansen, W. W., and J. G. Beckerley: Concerning New Methods of Calculating Radiation Resistance, Either with or without Ground, *Proc. IRE*, **24**, 1594, December, 1936.
34. ——— and J. R. Woodyard: A New Principle in Directional Antenna Design, *Proc. IRE*, March, 1938, p. 333.
35. Harrison, C. W., Jr., and R. King: Receiving Antenna in a Plane-polarized Field of Arbitrary Orientation, *Proc. IRE*, **32**, 35, January, 1941.
36. ———: Mutual Impedance of Antennas, *J. Applied Phys.*, **14**, 306, June, 1943.
37. ———: Approximate Representation of the Electromagnetic Field in the Vicinity of a Symmetrical Radiator, *J. Applied Phys.*, **15**, 544, July, 1944.
38. Howe, G. W. O.: Equivalent Inductance and Capacity of an Aerial with Inserted Tuning Coil or Condenser, *Exptl. Wireless*, **5**, 297, June, 1928.
39. ———: Applying Transmission Line Theory to Aerials, *Wireless Eng.*, January, 1939, p. 1.
40. ———: Calculation of Aerial Capacitance, *Wireless Eng.*, **20**, 157, April, 1943.
41. ———: Radiation Resistance of a Half-wave Dipole Aerial, *Wireless Eng.*, **22**, 153, April, 1945.
42. Jachnow, W.: Mutual Impedance of Inclined Rectilinear Conductors with Progressive Waves, *ENT*, July, 1939, p. 177.
43. ———: Theoretical Investigations of Radiation Diagrams and Radiation Resistance for Progressive Waves of Various Phase Velocities, abstract, *Wireless Eng.*, **20**, 87, February, 1943.
44. Jordan, E. C.: Acoustic Models of Radio Antennas, *Ohio State Univ. Eng. Theories*, **10** (3), May, 1941.
45. ——— and W. L. Everitt: Acoustic Models of Radio Antennas, *Proc. IRE*, April, 1941, p. 186.
46. Kelvin, Wm.: Radiation Field of an Unbalanced Dipole, *Proc. IRE*, **34**, 440, July, 1946.
47. King, R.: Approximate Representation of the Distant Field of Linear Radiators, *Proc. IRE*, **29**, 458, August, 1941.
48. ——— and F. G. Blake, Jr.: Self-impedance of a Symmetrical Antenna, *Proc. IRE*, **30**, 335, July, 1942.
49. ——— and C. W. Harrison, Jr.: Distribution of Current along a Symmetrical Center-driven Antenna, *Proc. IRE*, **31**, 548, October, 1943.
50. ———: Coupled Antennas and Transmission Lines, *Proc. IRE*, **31**, 626, November, 1943.
51. ——— and C. W. Harrison, Jr.: Receiving Antenna, *Proc. IRE*, **32**, 18, January, 1944.
52. ——— and ———: Impedance of Short, Long, and Capacitively Loaded Antennas with a Critical Discussion of the Antenna Problem, *J. Applied Phys.*, **15**, 170, February, 1944.
53. ——— and ———: Mutual and Self-impedance for Coupled Antennas, *J. Applied Phys.*, **15**, 481, June, 1944.
54. Labus, J.: Recherische Ermittlung der Impedanz von Antennen, *Hochfrequenztechn.*, p. 17, January, 1933. (Mutual impedance theory.)
55. Laport, E. A.: Calculation of Vertical Antenna Radiation Patterns, *Electronics*, February, 1935.
56. Levin, S. A., and C. J. Young: Field Distribution and Radiation Resistance of a Straight Vertical Unloaded Antenna Radiating at One of Its Harmonics, *Proc. IRE*, **14**, 675, October, 1926.
57. McPetrie, J. S.: Graphical Method for Determining the Magnitude and Phase of the Electric Field in the Neighborhood on an Antenna Carrying a Known Distribution of Current, *J. IEE*, **69**, 290, February, 1931.
58. ———: Method for Determining the Effect of the Earth on the Radiation from Aerial Systems, *J. IEE*, **70**, 382, March, 1932.
59. ——— and J. A. Saxon: Theory and Confirmation of Calibration of Field Strength Sets by Radiation, *J. IEE*, May, 1941, p. 11.
60. Moullin, E. B.: Radiation Resistance of Aerials Whose Length Is Comparable with the Wavelength, *J. IEE*, **78**, 540-563, May, 1936.
61. ———: Radiation Resistance of Surfaces of Revolution, Such as Cylinders, Spheres, and Cones *J. IEE*, **88**, 50, March, 1941.
62. Murray, F. H.: Mutual Impedance of Two Skew Antenna Wires, *Proc. IRE*, **21**, 154, January, 1933.
63. Neinan, M. S.: Principle of Reciprocity in Antenna Theory, *Proc. IRE*, **31**, 666, December, 1943.
64. Neissen, K. F.: Ground Absorption for Horizontal Dipole Aerials, *Ann. Physik*, July, 1938, p. 444.
65. ———: Choice between Horizontal and Vertical Dipole Aerials for Minimum Earth Absorption with Given Wavelength and Soil Type, *Ann. Physik*, November, 1938, p. 403.
66. ———: Radiation from a Dipole, *Ann. Physik*, February, 1937, p. 209.
67. ——— and G. deVries: Receiving Impedance of a Receiving Antenna, *Physica*, July, 1939, p. 601.
68. ———: Calculation of Field Strength of a Half-wave Aerial, *Physica*, September, 1940, p. 23.
69. Pistol Kors, A. A.: Radiation Resistance of Beam Antennas, *Proc. IRE*, **17**, 562, March, 1929.
70. Ramey, R. R.: Radiation and Induction, *Proc. IRE*, **21**, 1586, November, 1933.
71. Ratcliffe, J. A., L. G. Vedy, and A. F. Wilkins: Spreading of Electromagnetic Waves from a Hertzian Dipole, *J. IEE*, **70**, 522, May, 1932.

72. Schelkunoff, S. A.: On Diffraction and Radiation of E-m Waves, *Phys. Rev.*, August, 1939, p. 308.
73. ———: General Radiation Formula, *Proc. IRE*, October, 1939, p. 660.
74. ———: Theory of Antennas of Arbitrary Size and Shape, *Proc. IRE*, September, 1941, p. 493.
75. ——— and C. B. Feldman: Radiation from Antennas, *Proc. IRE*, 30, 511, November, 1942.
76. ———: Mathematical Theory of Linear Arrays, *Bell System Tech. J.*, 22, 80, January, 1943.
77. ———: Antenna Theory and Experiment, *J. Applied Phys.*, 18, 54, January, 1944.
78. Siegel, E. M.: Wavelength of Oscillations along Transmission Lines and Antennas, *Texas Univ. Eng. Research Ser.* 32, 1, 1940.
79. Southworth, G. C.: Certain Factors Affecting the Gain of Directive Antennas, *Proc. IRE*, 18, 1502, September, 1930.
80. Sommerfeld, A., and F. Renner: Radiation Energy and Earth Absorption for Dipole Antennae, *Wireless Engr.*, 19, 351, 409, 457, August–October, 1942.
81. Stratton, J. A., and L. J. Chu: Diffraction Theory of E-m Waves, *Phys. Rev.*, July, 1939, p. 99.
82. Thomson, W. T.: Development of the General Antenna Array Equation, *J. Applied Phys.*, 18, 420, May, 1944.
83. Walmsley, T.: Impedance Characteristics of Short-wave Dipoles, *Phil. Mag.*, June, 1938, p. 981.
84. Wells, N.: Aerial Characteristics, *J. IEE*, 89, Pt. 3, 76, June, 1942.
85. Wilmotte, R. M.: Distribution of Current in a Transmitting Antenna, *J. IEE*, 66, 617, June, 1928.
86. Wise, W. Howard: Grounded Condenser Antenna Radiation Formula, *Proc. IRE*, 19, 1684, September, 1931.
87. Zinke, C.: Fundamental Consideration of the Current and Potential Distribution on Antennas, *Wireless Engr.*, September, 1941, p. 377.

L-f Antenna Design:

88. Alexanderson, E. F. W.: Trans-Atlantic Radio Communication, *Proc. IRE*, August, 1920.
89. ———: Transoceanic Radio Communication, *Gen. Elec., Rev.*, October, 1920.
90. Ashbridge, N., H. Bishop, and B. N. MacLarty: Droitwich Broadcasting Station, *J. IEE*, 77, 447, 460; Discussion, 474, October, 1935.
91. Bailey, A., S. W. Dean, and W. T. Wintringham: Receiving System for Longwave Transatlantic Radiotelephony, *Bell System Tech. J.*, 8, 309, April, 1929.
- 91a. Beverage, H. H., C. W. Rice, and E. W. Kellogg: The Wave Antenna, *Trans. AIEE*, 42, 215, 1923.
92. Eckersley, T. L.: An Investigation of Transmitting Aerial Resistances, *J. IEE*, 60, 581–605, 1922. (Studies of direct earth and counterpoise capacity grounds.)
93. Kear, F. G.: Phase Synchronization in Directive Antenna Arrays with Particular Application to the Radio Range Beacon, *J. Research Nat. Bur. Standards*, 11, 123, July, 1933.
94. Lindenblad, N., and W. W. Brown: Main Considerations in Antenna Design, *Proc. IRE*, 14, 291, June, 1926.
95. Mann, H. F., and F. Hollinghurst: Replacement of Main Aerial System at Rugby Station, *POEE J.*, April 1940, p. 22.
96. Morgan, M. G.: Increasing Radiation at Low Frequencies, *Electronics*, July, 1940, p. 33.
97. Roseler, G., and K. Vogt: Investigations on Umbrella Aerials, abstract, *Wireless Engr.*, 20, 141, March, 1943.
98. Watson-Watt, R. A., and J. F. Herd: An Instantaneous Direction Reading, Radio Goniometer, *J. IEE*, 64, 611, 1926.

M-f Broadcast Antennas—Transmitting:

99. Andrew, V. J., and A. M. McGregor: Low Cost Wooden Mast; Station WJBC, *Electronics*, 9, 45, May, 1936.
100. Boudoux, P.: Current Distribution and Radiation Properties of the Shunt-excited Antenna, *Proc. IRE*, June, 1940, p. 271.
101. Brown, G. H.: Critical Study of the Characteristics of Broadcast Antennas as Affected by Antenna Current Distribution, *Proc. IRE*, 24, 48, January, 1936.
102. ———: Directional Antennas, *Proc. IRE*, 28, 78, January, 1937.
103. ——— and J. G. Leitch: Fading Characteristics of the Top-loaded WCAU Antenna, *Proc. IRE*, 25, 583, May, 1937.
104. ———, R. F. Lewis, and J. Epstein: Ground System as a Factor in Antenna Efficiency, *Proc. IRE*, 25, 753, June, 1937.
105. ———: A Radical Ground System Chart, *Electronics*, January, 1938, p. 33.
106. ———: WTAR Directional Array, *Electronics*, January, 1938, p. 38.
107. ———: Consideration of the Radio Frequency Voltages Encountered by the Insulating Materials of Broadcast Towers, *Proc. IRE*, September, 1939, p. 566.
108. ——— and J. M. Baldwin: Adjusting Unequal Tower Broadcast Arrays, *Electronics*, 16, 118, December, 1943.
109. Chamberlain, A. B., and W. B. Lodge: Broadcast Antenna, *Proc. IRE*, 24, 11, January, 1936.
110. Cox, C. R.: Phasing Networks for Broadcast Arrays; Graphical Methods Applied, *Electronics*, 17, 120, June, 1944.
111. Dewitt, J. H., Jr., and A. D. Ring: Significant Radiation from Directional Antennas of Broadcast Stations for Determining Sky-wave Interference at Short Distances, *Proc. IRE*, 32, 668, November, 1944.
112. Dutters, W. S.: Directional Antenna Chart, *Electronics*, February, 1940, p. 33.
113. ———: Design Chart for Phase Shifting and Amplitude Control Networks, *Electronics*, 15, 53, October, 1942.

114. ———: Adjustment of Directional Antennas, *Electronics*, **16**, 91, April, 1943.
 115. Eckersley, P. P., T. L. Eckersley, and H. L. Kirke: Design of Transmitting Aerials for Broadcasting Stations, *J. IEE*, **67**, 507, April, 1929.
 116. Eppen, F., and A. Gothe: Über die schwindvermindernde Antenne des Rundfunksenders Breslau, *ENT*, **10** (4), 1933. (Top-loaded vertical wire on wood tower for antifading.)
 117. Everest, F. A., and W. S. Pritchard: Horizontal Polar Pattern Calculator for Directional Broadcast Antennas, *Proc. IRE*, June, 1941, p. 355.
 118. "Rules of Good Engineering Practice for Standard Broadcast Stations," Federal Communications Commission, Washington, D. C.
 119. Feld, J.: Radio Antenna Suspended from 1,000 Foot Towers, *J. Franklin Inst.*, **239**, 363, May, 1945.
 120. Fitch, W. A., and W. S. Dutters: Measurement of Broadcast Coverage and Antenna Performance, *RCA Rev.*, July, 1938, p. 396.
 121. Gihring, H. E., and G. H. Brown: General Consideration of Tower Antennas for Broadcast Use, *Proc. IRE*, **23**, 311, April, 1935.
 122. Glas, E. T.: Efficiency-rating of Transmitting Aerials for Broadcasting Distribution, *Exptl. Wireless*, **7**, 665, December, 1930.
 123. Harbich, H., and W. Hahnemann: Wirksame Bekämpfung des Nahschwundes im Rundfunk durch Dendeantennengebilde bestimmter Form, *ENT*, **9** (10), October, 1932. (Treatment of circular arrays for antifading.)
 124. Harmon, R. N.: Some Comments on Broadcast Antennas, *Proc. IRE*, **24**, 36, January, 1936.
 125. Hutton, W. G., and R. M. Pierce: Mechanical Calculator for Directional Antenna Patterns, *Proc. IRE*, **30**, 233, May, 1942.
 126. Kear, F. G.: Maintaining the Directivity of Antenna Arrays, *Proc. IRE*, **22**, 847, July, 1934; **22**, 1313, November, 1934.
 127. Kirby, Samuel S.: Radio Field Intensity and Distance Characteristics of a High Vertical Broadcast Antenna, *Proc. IRE*, **24**, 859, June, 1936.
 128. Kleinwachter, H.: Determination of the Radiation Diagrams of Radiator Groups; Directive or Anti-fading Aerial Systems and Acoustic Radiators; an Electro-mechanical Method, abstract, *Wireless Engr.*, **19**, 576, December, 1942.
 129. Labus, J. W.: Broadcast Antenna for Low Angle Radiation, *Proc. IRE*, **23**, 935, August, 1935.
 130. Laport, E. A.: Directional Antenna Design, *Electronics*, **9**, 22, April, 1936.
 131. Leeman, A.: Modified Protective Gap for Transmitting Antennas, *Electronics*, **16**, 128, May, 1943.
 132. McKinnon, K. A.: Experimental Comparison of Shunt- and Series-excitation of a High Uniform Cross-section Vertical Radiator, *Can. J. Research*, November, 1939, p. 227.
 133. McPherson, W. L.: Electrical Properties of Aerials for Medium and Long Wave Broadcasting, *Elec. Commun.*, April, 1938, p. 306.
 134. Morrison, J. F., and P. H. Smith: Shunt-excited Antenna, *Proc. IRE*, **25**, 673, June, 1937.
 135. ———: Simple Method for Observing Current Amplitude and Phase Relations in Antenna Arrays, *Proc. IRE*, **25**, 1310-1326, October, 1937.
 136. ———: Simplifying Adjustments of Antenna Arrays, *Electronics*, January, 1940, p. 70.
 137. Prestholdt, Ogden: WABC's New Two-way Antenna, *Electronic Inds.*, **4**, 88, February, 1945.
 138. Roder, H.: Elimination of Phase Shifts between the Currents in Two Antennas, *Proc. IRE*, **22**, 374, March, 1934.
 139. Rountree, J. G.: Calculator for Two-element Directive Arrays, *Proc. IRE*, **32**, 760, December, 1944.
 140. Sinclair, D. B.: Impedance Measurements on Broadcast Antennas, *Communications*, **19**, 5, July, 1939.
 141. Smith, Carl E.: Critical Study of Two Broadcast Antennas, *Proc. IRE*, **24**, 1329, October, 1936.
 142. Smith, C. E., and E. L. Gove: Electromechanical Calculator for Directional Antenna Patterns, *Elec. Eng.*, **62**, *Trans.* 78, February, 1943.
 143. Spangenberg, K.: Charts for the Determination of the R.m.s. Value of Horizontal Radiation Pattern of Two-element Broadcast Arrays, *Proc. IRE*, **30**, 237, May, 1942.
 144. Taylor, J. P.: Cathode-ray Antenna Phasemeter, *Electronics*, **12**, 62, April, 1939.
 145. Williams, H. P.: Machine for Calculating Polar Diagrams, *Electronics*, **16**, 196, September, 1943.
- Broadcast, F-m, and Television Receiving Antennas:*
146. Acheves, J. G.: Antenna System for FM Reception, *Electronics*, September, 1941.
 147. Benham, W. E.: Aerial Coupling Systems for Television, *Wireless Engr.*, **15**, 555, October, 1938.
 148. Carlson, W. L., and V. D. Landon: New Antenna Kit Design, *RCA Rev.*, **2**, 60, July, 1937.
 149. ———: Television Reception with Built-in Antennas for Horizontally and Vertically Polarized Waves, *RCA Rev.*, **6**, 443, April, 1942.
 150. Fruin, D.: Central Antenna System, *Electronics*, November, 1939, p. 37.
 151. Goldman, Stanford: Shielded Loop for Noise Reduction in Broadcast Reception, *Electronics*, **2**, 20, October, 1938.
 152. Landon, V. D., and J. D. Reid: A New Antenna Design for Noise Reduction, *Proc. IRE*, March, 1939, p. 188.
 153. Morgan, H. K.: Multi-frequency Tuned Antenna System, *Electronics*, August, 1940, p. 42.
 154. Mountjoy, G.: Measurement of Effective Height of Automobile Antennas, *RCA Rev.*, **3**, 369, January, 1939.
 155. Noll, E. M.: Antennas for Television Receivers, *Radio News*, **33**, 40, May, 1945.
 156. Rettenmeyer, F. X.: Radio-frequency Distributing Systems, *Proc. IRE*, **23**, 1286, November, 1935.
 157. Seeley, S. W.: Effect of Receiving Antenna on Television Reception Fidelity, *RCA Rev.*, April, 1939, p. 433.

158. Sturley, K. A.: Receiver Aerial Coupling Circuits, *Wireless Engr.*, April, 1941, p. 137, and May, 1941, p. 190.
159. Taylor, S. G.: RCA Spiderweb Snags All Waves, *Radio News*, 18, 292, November, 1936.
160. Wheeler, H. A., and V. E. Whitman: Design of Doublet Antenna Systems for All-wave Radio Reception, *Proc. IRE*, 24, 1257, October, 1936.
161. New Form of Antenna; V-doublet System, *Electrician*, 114, 824, June 21, 1935. (Same abr. *Gen. Elec. Rev.*, 38, 395, August, 1935.)
162. Modern Antenna System for Apartment Houses, *Gen. Elec. Rev.*, May, 1939, p. 226.
163. Multicoupler Antenna System Includes FM Reception, *Rev. Sci. Instruments*, 13, 86, February, 1942.

H-f Antennas:

164. Barfield, R. H., and W. Ross: Short-wave Adecock Direction Finder, *J. IEE*, 81, 683, 1937; *Wireless Sec. IEE*, 13, 39, March, 1938.
165. Baumler, M., K. Kruger, H. Plendl, and W. Pfitzer: Radiation Measurements of a Short Wave Directive Antenna at the Nauen High Power Radio Station, *Proc. IRE*, May, 1931, p. 812.
166. Beverage, H. H., and H. O. Peterson: Diversity Receiving System of RCA Communications, Inc., for Radiotelegraphy, *Proc. IRE*, 19, 545, April, 1931.
167. Brown, R.: Transoceanic Radiotelephone Development; Multiple Unit Steerable Antenna, *Proc. IRE*, 25, 1131, September, 1937.
168. ———: Researches in Radiotelephony; Multiple-unit Steerable Antenna, *J. IEE*, 83, 395, September, 1939.
169. Bruce, E.: Developments in Short-wave Directive Antennas, *Proc. IRE*, 19, 1406, August, 1931; *Bell System Tech. J.*, 10, 656, October, 1931.
170. ———, A. C. Beck, and L. R. Lowry: Horizontal Rhombic Antennas, *Proc. IRE*, 23, 24, January, 1935.
171. ——— and ———: Experiments with Directivity Steering for Fading Reduction, *Proc. IRE*, 23, 357, April, 1935; *Bell System Tech. J.*, 14, 195, April, 1935.
172. Caferata, H.: A Generalized Radiation Formula for Horizontal Rhombic Antennas, *Marconi Rev.*, Nos. 80, 81, and 82; IX (1), January–March, 1946; IX (2), April–June, 1946; IX (3), July–September, 1946.
173. Carter, P. S., C. W. Hansell, and N. E. Lindenhlad: Development of Directive Transmitting Antennas by RCA Communications, *Proc. IRE*, 19, 1773, October, 1931.
174. Chamherlain, A. B.: CBS International Broadcast Facilities, *Proc. IRE*, 30, 118, March, 1942.
175. Everitt, W. L., and J. F. Byrne: Single Wire Transmission Lines for Short-wave Antennas, *Ohio State Univ., Eng. Exp. Sta. Bull.*, 52, 1, 1930.
176. Foster, Donald: Radiation from Rhombic Antennas, *Proc. IRE*, 25, 1327, October, 1937.
177. Foster, D. E., and C. W. Finnigan: Method of Measuring the Effectiveness of Electrostatic Loop Shielding, *Proc. IRE*, 31, 253, June, 1943.
178. Fereday, R. A.: Sense Finding Device for Use with Spaced-aerial Direction Finders, *J. IEE*, 84, 96, 1939.
179. Franz, K.: Gain and Absorption Area of Large Directional Aerials, *Hochfrequenztechn.*, December, 1939, p. 198.
180. Friis, H. T.: New Directional Receiving System, *Proc. IRE*, 13, 685, December, 1925.
181. ——— and C. R. Feldman: Multiple Unit Steerable Antenna for Short-wave Reception, *Proc. IRE*, 25, 841, July, 1937; *Bell System Tech. J.*, 16, 337, July, 1937.
182. Gardiner, E. L.: Aerial Reflectors, *Wireless World*, October, 1940, p. 419.
183. Gill, A. J.: Developments in Radio Engineering Carried out by the Post Office Engineering Department, *J. IEE*, February, 1939, p. 256.
184. Goddess, G.: NBC Beam Antennas, *Communications*, March, 1939, p. 16.
185. Goldman, S.: Dipoles and Reflectors, *Electronics*, May, 1940, p. 20.
186. Grosskopf, J., and K. Vogt: Polarization Measurements in the Field of Horizontal Transmitting Dipole, abstract, *Wireless Engr.*, 21, 437, September, 1944.
187. Hansen, W. W., and L. M. Hollingworth: Design of Flat-shooting Antenna Arrays, *Proc. IRE*, February, 1939, p. 137.
188. Harper, A. E.: "Rhombic Antenna Design," Van Nostrand, 1941.
189. Harrison, C. W. Jr.: Characteristics of the Two-antenna Array, *Proc. IRE*, 31, 75, February, 1943.
190. ———: Radiation from Vee Antennas, *Proc. IRE*, 31, 362, July, 1943.
191. ———: Radiation Field of Long Wires, with Application to Vee Antennas, *J. Applied Phys.*, 14, 537, October, 1943.
192. Hayes, L. W., and B. N. McLarty: Empire Service Broadcasting Station at Daventry, *J. IEE*, September, 1939, p. 328.
193. Howe, G. W. O.: Polar Diagram of a Simple Broadside Array, *Wireless Engr.*, 19, 193, May, 1942.
194. Kraus, J. D.: Multi-wire Dipole Antennas, *Electronics*, January, 1940, p. 26.
195. Lewin, L.: Rhombic Transmitting Aerial, *Wireless Engr.*, May, 1941, p. 140.
196. Meissner, A.: Directional Radiation with Horizontal Antennas, *Proc. IRE*, 15, 928, November, 1927.
197. Meisser, A., and H. Rothe: On the Determination of the Optimum Radiation Angle for Horizontal Antennas, *Proc. IRE*, 17, 35, January, 1929.
198. Page, H.: Measured Performance of Horizontal Dipole Transmitting Arrays, *Electrician*, 133, 408, Nov. 3, 1944; *Elec. Review (London)*, 135, 625, Nov. 3, 1944.

199. Polkinghorn, F. A.: Single Sideband MUSA Receiving System for Commercial Operation on Transatlantic Radiotelephone Circuits, *Bell System Tech. J.*, April, 1940, p. 306.
200. Ross, W.: Calibration of 4-aerial Adcock Direction Finders, *J. IEE*, August, 1939, p. 192.
201. Siegel, E.: Reception with Short Antennas, *Electronics*, 9, 48, January, 1936.
202. Smale, J. A.: Comparative Merits of Different Types of Directive Aerials for Communications, *J. IEE*, 91, 3, 12, March, 1944; *Electrician*, 132, 74, Jan. 28, 1944; *Elec. Review (London)*, 134, 119, Jan. 28, 1944.
203. Sterba, E. J.: Theoretical and Practical Aspects of Directional Transmitting Systems, *Proc. IRE*, 19, 1184, July, 1931.
204. Van Slooten, J.: Receiving Aerials, *Philips Tech. Rev.*, November, 1939, p. 320.
205. Walmsley, T.: Beam Arrays and Transmission Lines, *J. IEE*, February, 1931.
206. ———: New Type of Directive Aerial, *Wireless Engr.*, 9, 662, November, 1932.
207. Wells, N.: Beam Wireless Telegraphy, *Elec. Review (London)*, 102, 898, 940, May 25–June 1, 1928.
208. Wells, N.: Short-wave Dipole Aerials, *Wireless Engr.*, 20, 219, May, 1943.
209. ———: Quadrant Aerial; an Omni-directional Wide-band Horizontal Aerial for Short Waves, *J. IEE*, 91, Part 3, 182, December, 1944; Brief paper, *Marconi Rev.*, 9, January–March, 1946
210. Chart for Rhombic Antenna Design, *Electronic Inds.*, 3, 112, October, 1944.
211. Marconi Series-phase Aerial, *Engineering*, 139, 235, Aug. 31, 1934.

V-h-f Antennas:

212. Alford, A., and A. G. Kandoian: UHF Loop Antennas, *Elec. Eng.*, 59, 843–848, April, 1940; *Elec. Commun.*, 18, 255–265, April, 1940.
213. ———: Ultrashort Electromagnetic Waves; Design of Radiators or Antennas for Ultrashort Waves, *Elec. Eng.*, 62, 303, 338, July–August, 1943.
214. Barfield, R. H., and R. A. Fereday: Improved Medium Wave Adcock Direction Finder, *Wireless Sec.*, *IEE*, 13, 33, March, 1938.
215. Barrow, W. L., and F. M. Greene: Rectangular Hollow Pipe Radiators, *Proc. IRE*, December 1938, p. 1498.
216. Barrow, W. L., and L. J. Chu: Theory of the Electromagnetic Horn, *Proc. IRE*, January, 1939, p. 51.
217. ——— and F. D. Lewis: Sectoral Electromagnetic Horn, *Proc. IRE*, January, 1939, p. 41
218. ———, L. J. Chu, and Jansen: Biconical Electromagnetic Horns, *Proc. IRE*, December, 1939, p. 769.
219. ——— and C. Shulman: Multi-unit Electromagnetic Horns, *Proc. IRE*, March, 1940, p. 130.
220. Bromel, R.: Beam Properties of Small Parabolic Reflectors with Various Excitations, *Hochfrequenz techn.*, 48, 81, 121, 1936.
221. Brown, G. H.: Turnstile Antenna for UHF Broadcasting, *Electronics*, April, 1936.
222. ——— and J. Epstein: UHF Antenna of Simple Design, *Communications*, July, 1940, p. 3.
223. ——— and ———: Pretuned Turnstile Antenna, *Electronics*, 18, 102, June, 1945.
224. Builder, G.: Duplex Operation on a Single Aerial, *Electronics*, February, 1940, p. 61.
225. Caporale, Peter: CAA Instrument Landing System, *Electronics*, March, 1945 (excellent bibliography on blind landing systems); February, 1945, p. 116, and March, 1945 p. 128.
226. ———: CAA-RTCA Instrument Landing Installation, *Electronic Inds.*, March, 1944, p. 92.
227. Carter, P. S.: Simple Television Antennas, *RCA Rev.*, October, 1939, p. 168.
228. ———: Antenna Arrays around Cylinders, *Proc. IRE*, 31, 671, December, 1943.
229. Chu, L. J., and W. L. Barrow: E-m Waves in Hollow Metal Tubes of Rectangular Cross-section, *Proc. IRE*, December, 1938, p. 1520.
230. ——— and ———: Electromagnetic Horn Design, *Elec. Eng.*, July, 1939, p. 333
231. ———: Calculation of Radiation Properties of Hollow Pipes and Horns, *J. Applied Phys.*, September, 1940, p. 603.
232. Dolph, C. L.: A Current Distribution for Broadside Arrays Which Optimizes the Relationship between Beamwidth and Side-lobe Level, *Proc. IRE*, 34, 335, June, 1946.
233. Foster, Donald: Loop Antennas with Uniform Current, *Proc. IRE*, October, 1944, pp. 603–607.
234. Franklin, J. C.: Installation of Antennas for Direction Finding, *Air Commerce Bull.* 9, 256, April, 1938.
235. Hasenbeck, H. W.: Design Data for Ground Plane Antennas, *Electronics*, 16, 98, August, 1943.
236. Hansell, C. W.: Radio Relay Systems Developed by the Radio Corporation of America; Couplings between Directional Antennas, *Proc. IRE*, 33, 164, March, 1945.
237. Herold, E. W., and L. Malter: Some Aspects of Radio Reception at Ultra-high Frequency; the Antenna and the Receiver Input Circuits, *Proc. IRE*, 31, 423, August, 1943.
238. Hromada, J. C.: Ultra-high Frequency Antennas in Aviation Radio, *Radio News*, 27, 14, May, 1942.
239. Kandoian, A. G.: Radiating Systems and Wave Propagation, *Electronics*, 15, 39, April, 1942.
240. ———: 3 New Antenna Types and Their Applications, *Elec. Commun.*, March, 1946, pp. 27–34; also *Proc. IRE*, 34, 70W–75W, February, 1946.
241. Kenyon, R. W.: Principles of Loran in Position Location, *Electronic Inds.*, December, 1945, p. 106.
242. ———: Loran Transmitting Stations, *Electronics*, March, 1946, p. 109.
243. Kraus, J. D.: Antenna Arrays with Close-spaced Elements, *Proc. IRE*, February, 1940, p. 76.
244. ———: Corner-reflector Antenna, *Proc. IRE*, November, 1940, p. 513.
245. Lange, P.: Measurements on Dipoles in the Decimetric Region, *Wireless Engr.*, November, 1941, p. 465.

246. Laport, Edmund A., and James B. Knox: Radiating System for 75 Megacycle Cone-of-silence Marker, *Proc. IRE*, **30**, 26, January, 1942.
247. Lesensky, M. M.: UHF Aerial Coupling Circuits, *Indian^{*}J. Phys.*, December, 1940, p. 67.
248. Lewis, R. F.: Vertical Turnstile Antenna, U.S. Patent 2210491.
249. Lindenblad, N. E.: Television Transmitting Antenna for Empire State Bldg., *RCA Rev.*, April, 1939, p. 387.
250. McPherson, W. L., and E. H. Ulrich: Micro-ray Communication, *J. IEE*, **78**, 632, June, 1936.
251. Morgan, H. K.: Beacon Antenna Characteristics, *Air Commerce Bull.*, **9**, 77, October, 1937.
252. Moullin, E. B.: Theory and Performance of Corner Reflectors for Aerials; abstracts, *Electrician*, **133**, 408, Nov. 3, 1944; *Elec. Review (London)*, **135**, 625, Nov. 3, 1944.
253. Nagy, A. W.: Experimental Study of Parasitic Wire Reflectors on 2.5 Meters, *Proc. IRE*, **24**, 233, February, 1936.
254. Naschke, E. S.: Police Satellite System; FM Signals Automatically Relayed to Central Point by Two Mountain-top Stations; Design of 60-degree Corner Reflector Receiving Antennas, *Electronica*, **17**, 94, May, 1944.
255. Palmer, L. S., and K. G. Gillard: Distribution of Ultra-high-frequency Currents in Long Transmitting and Receiving Antennae, *J. IEE*, **83**, 415, September, 1938.
256. Parker, R. H.: Coaxial Antenna for 112 Mc., *QST*, **29**, 40, June, 1945.
257. Polydroff, W. J.: Iron-core Loop Antennas, *Electronic Inds.*, **3**, 84, November, 1943.
258. Reber, G.: Electromagnetic Horns, *Communications*, February, 1939, p. 13.
259. Schafer, O.: Comparison between Electric Horns and Other Directional Radiators, *Wireless Engr.*, **22**, 90, February, 1945.
260. Scheldorf, M. W.: FM circular antenna, *Gen. Elec. Rev.*, **46**, 163, March, 1943.
261. ———: New Studio-to-transmitter Antenna, *Proc. IRE*, **33**, 106, February, 1945.
262. Seeley, S. W.: Shoran Precision Radar, *Elec. Engr.*, April, 1946, p. 232.
263. Sherman, J. B.: Circular Loop Antennas at Ultra-high Frequencies, *Proc. IRE*, **32**, 534, September, 1944.
264. Smith-Rose, R. L., and H. G. Hopkins: Radio Direction Finding on Wavelengths between 6 and 10 Meters, *J. IEE*, **83**, 87, 1938; *Wireless Sec.*, *IEE*, **13**, 303, September, 1938.
265. ——— and H. G. Hopkins: Radio Direction-finding on Wavelengths between 2 and 3 Meters, *J. IEE*, August, 1940, p. 154.
266. Southworth, G. C., and A. P. King: Metal Horns as Directive Receivers of Ultra-short Waves, *Proc. IRE*, February, 1939, p. 95.
267. Stahl, C. I. H. A.: Full Parabolic Reflectors for Microwaves, *Philips Transmitting News*, **3**, 14, 1937.
268. Taylor, J. P.: Square-loop F-m Antenna at WBRL, *Electronics*, **18**, 96, March, 1945.
269. Uda, S.: High Angle Radiation of Short Electric Waves, *Proc. IRE*, **15**, 377, May, 1927.
270. Williams, E. M.: Radiation Characteristics of Short-wave Loop Aerials, *Proc. IRE*, October, 1940, p. 480.
271. Wolff, Irving: Determination of the Radiating System Which Will Produce a Specified Directional Characteristic, *Proc. IRE*, **25**, 630, May, 1937.

Radar Antennas:

272. AN/TPS-3, 590-610 Mc, *Electronics*, February, 1946, pp. 98-100; January, 1946, pp. 98-100.
273. CXAS 8 Elements with Cylindrical Parabola, 680-720 Mc, *Bell System Tech. J.*, January, 1946, pp. 7-9.
274. USN Marks 1, 2, 3,—8 Elements with Cyl. Par. Refl., *Bell System Tech. J.*, January, 1946, p. 23.
275. USN Mark 4 Double—8 Elements with Cyl. Par. Refl., *Bell System Tech. J.*, 1946, pp. 40-43.
276. AN/MPG-1 (Bath-tub Radiator) 10,000 Mc, *Electronics*, January, 1946, pp. 110-117.
277. SCR-584 (Par. Refl. with Dipole and Disc), *Electronics*, December, 1945, p. 104.
278. SCR-268 (Lobe-switched Flat Array), *Electronics*, September, 1945, p. 100.
279. AAF Microwave Blind Landing System, *Electronic Inds.*, February, 1946, pp. 60-64.
280. Radio Link; Diversity Reception, *Electrician*, **119**, 573, Nov. 12, 1937.
281. Pennsylvania Turnpike U-h-f Traffic Control System; Antenna and Feeder Details, *Electronics*, **15**, 40, May, 1942.
282. New FM Radio Relay Antenna, *Electronics*, **16**, 230, June, 1943.
283. Circular Antenna for Ultrahigh Frequencies, *QST*, **26**, 19, November, 1942.
284. Antennas for 112 Mc. Mobile Work, *QST*, **26**, 14, February, 1942.

Aircraft Antennas:

285. Haller, G. L.: Constants of Fixed Antenna on Aircraft, *Proc. IRE*, April, 1938, p. 415.
286. Levy, G. F.: Design Problems of Directional Loop Antennas for Aircraft, *Aviation*, **41**, 135, March, 1942.

Addenda:

287. Holts, R. F.: The Pylon Antenna, *Broadcast News*, **44**, 36, October, 1946.
288. Brown, G. H., and Morrison: The Antennalyzer, *Broadcast News*, **43**, 6, June, 1946.
289. Kock, W. E.: Metal Lens Antennas, *Proc. IRE*, **34** (11), 828-836, November, 1946.
290. Brown, G. H., and Woodward: Circularly-polarized Omnidirectional Antenna, *RCA Rev.*, **3**, 259, June, 1947.
291. Friis, H. T., and Lewis: Radar Antennas, *Bell System Tech. J.*, **26**, 219, April, 1947.

CHAPTER 15

WAVE GUIDES AND CAVITY RESONATORS

BY THEODORE MORENO¹

1. **Relationship between Wave Guides and Transmission Lines.** A transmission line may be defined as a system of material boundaries forming a continuous path from one place to another and capable of directing energy along this path. The electromagnetic energy is carried along a transmission line in the form of guided electromagnetic waves. Transmission-line theory may be developed in terms of these traveling waves which propagate in either direction along the line.

Wave guides are included as transmission lines by the above definition. This is justified because many wave-guide phenomena are readily interpreted in terms of the traveling-wave theory of transmission lines, and it is possible to draw equivalent transmission-line circuits for many wave-guide structures. Certain modifications to ordinary line theory are required for wave-guide transmission circuits. To determine

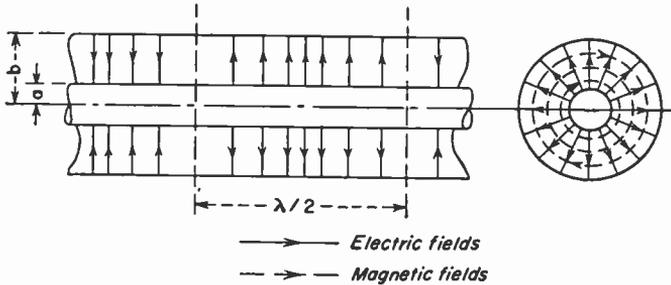


FIG. 1. Electric and magnetic fields for principal mode of propagation in coaxial transmission line.

these modifications, it is necessary to examine the nature of traveling waves on transmission lines.

A transmission line at ordinary radio frequencies consists of two or more separated conductors. When the line is energized, currents flow in these conductors and voltages exist between conductors. Electric and magnetic fields are found in the insulating medium separating the conductors. An example is the coaxial line shown in Fig. 1. The electric lines of force are shown as solid lines, and the magnetic lines of force as broken lines. The electric and magnetic lines of force are at right angles to each other, as is usually characteristic of electromagnetic waves. In the example, the electric and magnetic lines of force are also everywhere transverse to the direction of energy propagation, which is along the axis of the conductors. For this reason, the wave shown is called a *transverse electromagnetic wave*, abbreviated *TEM*.

TEM waves are used in all conventional transmission lines. If there are only two conductors, only one *TEM* wave is possible. If there are more than two conductors, more than one *TEM* wave may exist.²

¹ Research Laboratory of Electronics, Massachusetts Institute of Technology.

² For example, in a two-wire overground transmission line, there is one wave in which the current flows down one wire and returns in the other, and a second wave in which the current flows down both wires and returns through the ground.

There may exist on a transmission line, in addition to the limited number of *TEM* waves, an infinite number of other types of waves, or higher modes of transmission. Each of these other modes has its own distinctive configuration of electric and magnetic fields. For a coaxial line, a few of the higher modes are shown in Fig. 2. The *TEM* wave, also called the *principal mode*, is distinguished from these higher modes in a number of ways. For example:

1. The transmission line acts as a high-pass filter for the higher modes, and the line will carry energy only if the frequency is above a cutoff frequency. This cutoff frequency is in general different for each higher mode. The higher modes may be excited at frequencies lower than their cutoff frequencies, but their strength will diminish very rapidly with increasing distance from the point of excitation.
2. The wavelength measured on the transmission line will be greater for the higher modes than for the principal mode.
3. The characteristic impedance and propagation constant of the line for the higher modes will differ from the values for the principal mode. The characteristic impedance is harder to define for the higher modes, as there no longer exists a unique value of voltage between conductors.

A transmission line requires two separated conductors to support a principal mode, but higher modes can carry energy on transmission lines that have only one conductor. For example, the coaxial line modes shown in Fig. 2 will reduce to the circular waveguide modes shown in Fig. 7 as the diameter of the center conductor is reduced to zero.

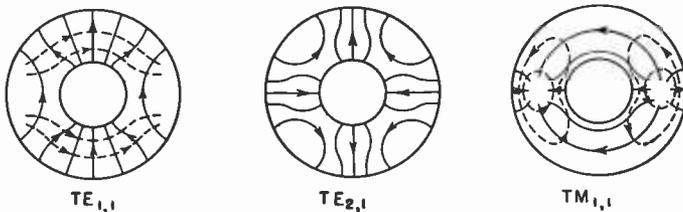


FIG. 2. Field configurations for some of the higher modes on coaxial line.

A wave guide is a transmission line not operating in a *TEM* mode. Most wave guides are incapable of supporting a *TEM* wave. They must therefore have less than two separated conductors. Normally a wave guide consists of a hollow conducting cylinder, of arbitrary cross section, with the electromagnetic fields confined to the inside of the cylinder. A dielectric rod may also act as a wave guide in the absence of any conductor.

Because of the phenomenon of cutoff that is peculiar to wave guides, the required dimensions of a wave guide for energy transmission are larger at lower frequencies. For most applications, wave guides are impractically large and bulky at frequencies below 1,000 or 2,000 Mc. On the other hand, conventional transmission lines, such as coaxial lines, must be restricted in size to keep the higher modes below cutoff¹ and become impractically small at frequencies much higher than 10,000 Mc. In the region between 1,000 and 10,000 Mc, both wave guides and coaxial lines are extensively used. Coaxial lines are smaller and lighter and, if cables are used, more flexible. On the other hand, wave guides offer lower attenuation, greater power-carrying capacity, and mechanical simplicity. A rectangular wave guide has a practical frequency range of less than two to one, while the band width of a coaxial line is much greater. A comparison of a typical wave guide, air dielectric coaxial line, and flexible cable is given in Table 1 for a frequency of 5,000 Mc.²

¹ In a coaxial line, the cutoff wavelength of the first of the higher modes is given approximately by

$$\lambda_c = \sqrt{\epsilon} \pi (b + a)$$

where *b* and *a* are the outer and inner radii of the line and ϵ is the dielectric constant of the medium separating the conductors.

² Moreno, T., Engineering Approach to Wave Guides, *Electronics*, 19 (5), 99, May, 1946.

Table 1. Comparison of Transmission Lines for 5,000 Mc

Type of line	Rectangular wave guide	Rigid coaxial line	Flexible cable
Army-Navy type No.	RG-49/U	RG-76/U	RG-9/U
Outside dimensions.	2 × 1 in.	5/8 in. diam	0.420 in. diam
Conductor material.	Brass	Brass	Inner conductor, 7/21 AWG silvered copper Outer conductor, double braid, inner silvered copper, outer copper
Surface finish.	Silver	Silver	Silver
Dielectric.	Air	Air	Polyethylene
Weight, lb per ft.	1.40	0.292	0.150
Attenuation, db per ft.	0.011	0.035	0.23
Recommended power rating	1.2 Mw	0.3 Mw	4,000 volts rms max, 66 watts continuous

2. Characteristics of Wave-guide Modes.¹ The higher modes which may carry energy in uniform wave guides with homogenous dielectric are divisible into two classes: *TE* and *TM* modes, each infinite in number for all wave guides.

TE (Transverse Electric) Modes. The electric field is everywhere transverse to the direction of energy propagation, which is the direction of the guide axis. The magnetic field has transverse components, but in addition has components parallel to the axis. These modes are also called *H* waves.

TM (Transverse Magnetic) Modes. The magnetic field is everywhere transverse to the guide axis. The electric field has transverse components, but in addition has components parallel to the axis. These modes are also called *E* waves.

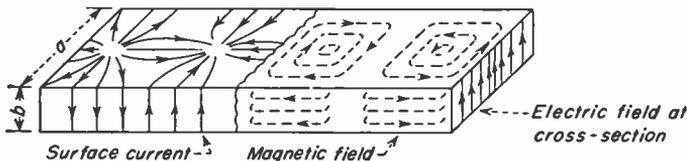


FIG. 3. Dominant mode, $TE_{1,0}$, in rectangular wave guide, showing electric and magnetic fields and conduction currents in walls.

It is customary to identify individual modes by an additional subscript notation, the rules of which depend upon the guide configuration. For circular and rectangular cross sections, these rules are discussed in the following paragraphs.

For any wave guide, the mode of transmission that has the lowest cutoff frequency is called the *dominant mode*.

*Rectangular Wave Guides.*² Individual modes in rectangular wave guides are identified by giving the class of the transmission mode, followed by two numerical subscripts. For example, the dominant mode in rectangular wave guide, illustrated in Fig. 3, is the $TE_{1,0}$ mode. The first and second subscripts indicate the number of half-period variations in transverse field intensity along the larger and smaller cross-section dimensions, denoted by a and b , respectively. A few of the higher modes of transmission are illustrated in Fig. 4. The $TE_{1,0}$ mode in rectangular wave guide deserves special consideration, because it is by far the most widely used mode of transmission in practical wave guides. Also, a simple physical picture of the wave

¹ The four articles that form the basis of modern development of wave guides are: Southworth, G. C., *Hyper-frequency Wave Guides—General Considerations and Experimental Results*, *Bell System Tech. J.*, 15 (2), 284-309, April, 1936. Carson, J. R., S. P. Mead, and S. A. Schelkunoff, *Hyper-frequency Wave Guides—Mathematical Theory*, *Bell System Tech. J.*, 15 (2), 310-313, April, 1936. Barrow, W. L., *Transmission of Electromagnetic Waves in Hollow Tubes of Metal*, *Proc. IRE*, 24 (10), 1298, October, 1936. Chu, L. J., and W. L. Barrow, *Electromagnetic Waves in Hollow Metal Tubes of Rectangular Cross Section*, *Proc. IRE*, 26 (12), 1520, December, 1938.

² Chu and Barrow, *loc. cit.*

structure is possible, from which many of the basic formulas that apply to all types of wave guides may be developed.

The fields inside the wave guide may be considered to result from the addition of two plane waves, whose electric fields are parallel to the side walls of the guide.¹ Each of these waves advances at an angle with respect to the axis of the guide, and each is the reflection of the other from the side walls of the guide. For the tangential electric field to be zero at the conducting side walls, the two plane waves must be out of phase at the walls; *i.e.*, the positive crest of one wave coincides with the negative crest of the other. This is illustrated in Fig. 5. Maximum electric field intensities are found at the center of the guide, where the two traveling waves are in phase and reinforce each other.

When the frequency is much higher than the cutoff frequency, the two component plane waves travel nearly parallel to the axis of the guide. As the frequency decreases toward cutoff, the angle between the guide axis and the direction of propagation of

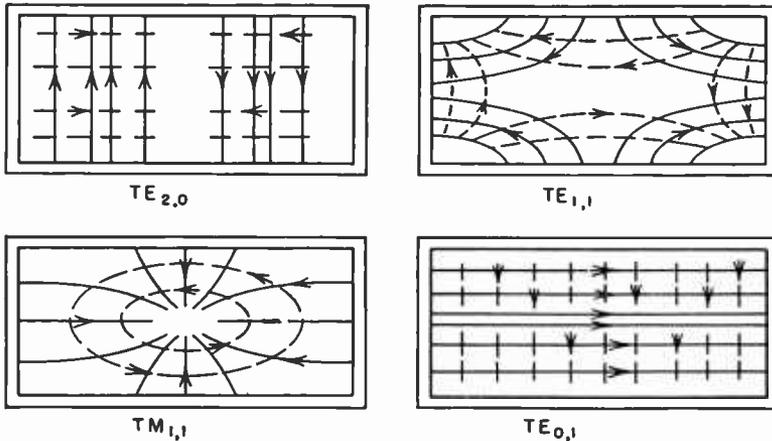


FIG. 4. Some of the higher modes of transmission in rectangular wave guide.

the component waves increases. At cutoff, this angle reaches 90 deg, and the two component waves are reflected back and forth between the side walls with no propagation of energy down the guide. At cutoff, the distance between the side walls is one-half a free-space wavelength. For the $TE_{1,0}$ mode in rectangular wave guide, the cutoff wavelength λ_c is related to the larger dimension of the wave guide a by

$$\lambda_c = 2a \tag{1}$$

The distance between positive and negative crests measured along the axis of the guide is seen by Fig. 5 to be greater than the distance between positive and negative crests of the component plane waves. This leads to the very important conclusion that the apparent wavelength in a wave guide, which is the distance between surfaces of equivalent phase, is greater than the free-space wavelength at the same frequency. If a wave guide is used as a Lecher wire system to measure wavelength, this correction must be made.

More complex modes in rectangular wave guide may be constructed by the superposition of more than two plane waves. The method is applicable, in fact, to wave guides of any cross section, although in general an infinite number of plane waves are required and the calculation becomes tedious.²

¹ Page, L., and N. I. Adams, Jr., *Electromagnetic Waves in Conducting Tubes*, *Phys. Rev.*, **52** (18), Sept. 15, 1937. Also see H. H. Skilling, *Physical Behavior of Wave Guides*, *Electronics*, **16** (3), 76, March, 1943, which gives a less mathematical but more graphical development.

² Mackeown, S. S., and J. W. Miles, *The Plane Wave Resolution of Guided Waves*, *Proc. IRE*, **33** (11), 805, November, 1945.

3. Phase and Group Velocity. The component plane waves, which add to give the fields inside the guide, travel at the speed of light (3×10^{10} cm per sec). Since the wave-guide wavelength λ_g is greater than the wavelength of the component waves λ , this gives rise to a *phase velocity* in the wave guide v_p , which exceeds the velocity of light c by the ratio v_p/c . The rate of energy propagation down the guide is not equal to the phase velocity, however, but is instead less than the velocity of light, because

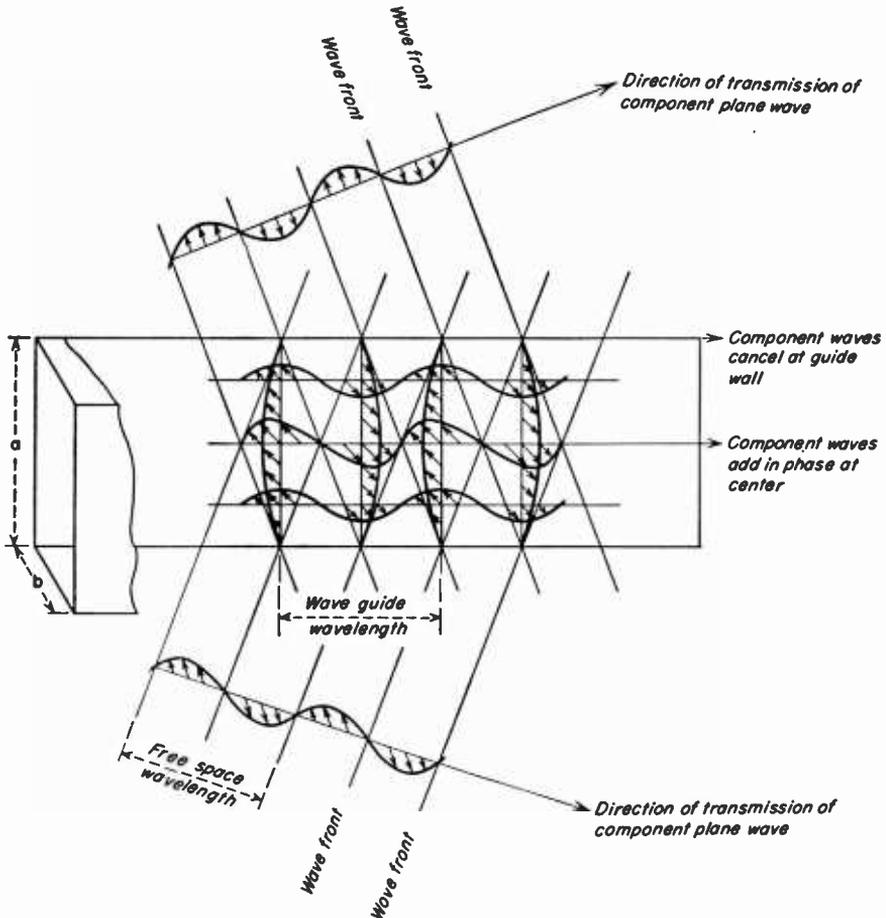


FIG. 5. Construction of the dominant mode in rectangular waveguide by the superposition of two plane waves traveling at an angle with respect to guide walls.

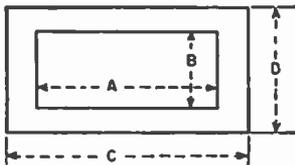
the component waves which carry the energy travel at an angle with respect to the guide axis. The rate of energy propagation along the axis is equal to the projection upon the axis of the velocity vector of the component waves. This rate of energy propagation is known as the *group velocity* v_g . The group velocity and the phase velocity are related to the velocity of light by

$$v_p v_g = c^2 \quad (2)$$

This relationship holds for any mode in any air-filled wave guide.

Table 2. Dimensions, Tolerances, and Frequency Range for Rectangular Wave Guides

RMA designation	Frequency range (1,000 Mc), for dominant (TE_{10}) mode		Dimensions, in.							Max inner radius†	
			Inner			Outer			Wall thickness		
			A	B	Tolerance	C	D	Tolerance	Nominal		Deviation from mean
WR770		0.96-1.45	7.700	3.850	± .005	7.950	4.100	± .005	0.125	± .005	3/64
WR650*	1.12-1.70		6.500	3.250	± .005	6.660	3.410	± .005	0.080	± .005	3/64
WR510		1.45-2.20	5.100	2.550	± .005	5.260	2.710	± .005	0.080	± .005	3/64
WR430*	1.70-2.60		4.300	2.150	± .005	4.460	2.310	± .005	0.080	± .005	3/64
WR340*		2.20-3.30	3.400	1.700	± .005	3.560	1.860	± .005	0.080	± .005	3/64
WR284*	2.60-3.95		2.840	1.340	± .005	3.000	1.500	± .005	0.080	± .005	3/64
WR229		3.30-4.90	2.290	1.145	± .005	2.418	1.273	± .005	0.064	± .005	3/64
WR187*	3.95-5.85		1.872	0.872	± .005	2.000	1.000	± .005	0.064	± .005	3/32
WR159		4.90-7.05	1.590	0.795	± .004	1.718	0.923	± .004	0.064	± .005	3/32
WR137*	5.85-8.20		1.372	0.622	± .004	1.500	0.750	± .004	0.064	± .005	3/32
WR112*		7.05-10.00	1.122	0.497	± .004	1.250	0.625	± .004	0.064	± .005	3/32
WR90*	8.20-12.40		0.900	0.400	± .003	1.000	0.500	± .003	0.050	± .005	3/32
WR75		10.00-15.00	0.750	0.375	± .003	0.850	0.475	± .003	0.050	± .005	3/32
WR62*	12.40-18.00		0.622	0.311	± .0025	0.702	0.391	± .003	0.040	± .005	3/64
WR51		15.00-22.00	0.510	0.255	± .0025	0.590	0.335	± .003	0.040	± .005	3/64
WR42*	18.00-26.50		0.420	0.170	± .0020	0.500	0.250	± .003	0.040	± .005	3/64
WR34		22.00-33.00	0.340	0.170	± .0020	0.420	0.250	± .003	0.040	± .005	3/64
WR28*	26.50-40.00		0.280	0.140	± .0015	0.360	0.220	± .002	0.040	± .003	3/64
WR22*		33.00-50.00	0.224	0.112	± .0010	0.304	0.192	± .002	0.040	± .003	0.010
WR19	40.00-60.00		0.188	0.094	± .0010	0.268	0.174	± .002	0.040	± .003	0.010
WR15*		50.00-75.00	0.148	0.074	± .0010	0.228	0.154	± .002	0.040	± .003	0.008
WR12*	60.00-90.00		0.122	0.061	± .0005	0.202	0.141	± .002	0.040	± .003	0.006
WR10		75.00-110.00	0.100	0.050	± .0005	0.180	0.130	± .002	0.040	± .003	0.006



* Military standard.
 † For all sizes: Min outer radius 3/64 in. Max outer radius 3/32 in.

The relationship between free-space wavelength λ , wave-guide wavelength λ_g , and cutoff wavelength λ_c for any mode in any air-filled wave guide is given by

$$\lambda_g = \frac{\lambda}{\sqrt{1 - (\lambda/\lambda_c)^2}} \tag{3}$$

For the important $TE_{1,0}$ mode in rectangular wave guide, the cutoff wavelength is given by Eq. (1), and Eq. (3) reduces to

$$\lambda_g = \frac{\lambda}{\sqrt{1 - (\lambda/2a)^2}} \tag{4}$$

For all TE and TM modes in rectangular wave guide, the cutoff wavelength is given by

$$\lambda_c = \frac{2}{\sqrt{\left(\frac{m}{a}\right)^2 + \left(\frac{n}{b}\right)^2}} \quad (5)$$

In Eq. (5), m and n are the subscripts denoting the particular mode under consideration (e.g., $TE_{m,n}$), and also indicate the number of half-period variations in transverse field intensity along the a and b cross-section dimensions of the guide, respectively.

If a rectangular guide is to be above cutoff for the dominant $TE_{1,0}$ mode, but below cutoff for all other modes,

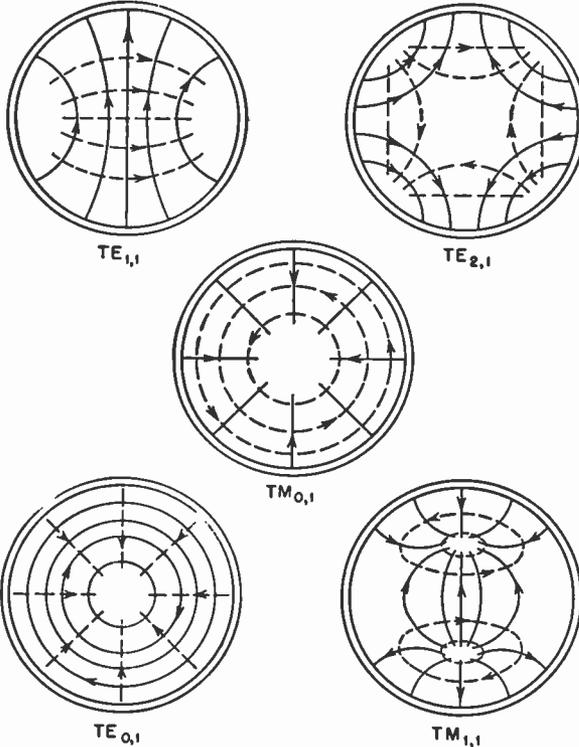


FIG. 6. Some transmission modes in circular wave guides.

with the wall, and both vary in a manner related to a Bessel function of the first kind of order m along a radius. Any particular mode is identified by the notation $TE_{m,n}$ or $TM_{m,n}$ where m is the total number of full-period variations of either component of field along a circular path concentric with the wall, and n is one more than the total number of reversals of sign of either component of field along a radial path.

The dominant mode in circular wave guide, that with the longest cutoff wavelength, is the $TE_{1,1}$ mode, which corresponds to the $TE_{1,0}$ mode in rectangular guide. This mode is shown in Fig. 6. Another mode that is of considerable interest is the $TM_{0,1}$, or circular magnetic mode. This is the lowest mode with circular symmetry. A mode of considerable theoretical interest is the $TE_{0,1}$ mode, which has anomalous attenuation characteristics discussed below.

The cutoff wavelength for all modes in circular wave guide depends upon the ratio of diameter to wavelength. For the $TE_{m,n}$ modes, the cutoff wavelength is given by

¹ Southworth, *loc. cit.* Carson, Mead, and Schelkunoff, *loc. cit.* Barrow, *loc. cit.*

$$a > \frac{\lambda}{2} < \lambda$$

$$b < \frac{\lambda}{2}$$

Various standard sizes of rectangular wave guide for board-band use have been recommended by the RMA Subcommittee on Microwave Transmission Lines. These are given in Table 2, taken from *Electronics*, p. 111, June, 1949, along with the frequency range for which each of the guides is recommended.

4. Circular Wave Guides.¹ A number of possible modes of transmission in circular wave guides are shown in Fig. 6. For any mode of transmission in a circular wave guide, the transverse fields may be resolved into two components: tangential and radial. Both of these components vary periodically along a circular path concentric

$$\lambda_c = \frac{2\pi a}{u_{m,n}'} \tag{6}$$

where a is the inner radius of the guide. The constant $u_{m,n}'$ is the n th root of the equation $J_m'(u) = 0$, where the prime denotes differentiation. Some of the lower values of $u_{m,n}'$ are

$u_{01}' = 3.832$	$u_{02}' = 7.016$	$u_{12}' = 8.536$
$u_{11}' = 1.841$	$u_{12}' = 5.332$	
$u_{21}' = 3.054$	$u_{22}' = 6.706$	
$u_{31}' = 4.201$	$u_{32}' = 8.016$	

For the $TM_{m,n}$ modes, the cutoff wavelength is given by

$$\lambda_c = \frac{2\pi a}{u_{m,n}} \tag{7}$$

where a is the inner radius of the guide and $u_{m,n}$ is the n th root of the equation $J_m(u) = 0$. Some of the lower values of $u_{m,n}$ are

$u_{01} = 2.405$	$u_{02} = 5.520$	$u_{03} = 8.654$
$u_{11} = 3.832$	$u_{12} = 7.016$	
$u_{21} = 5.136$	$u_{22} = 8.417$	
$u_{31} = 6.380$		

The cutoff frequency of the $TM_{0,1}$ mode is only 30 per cent higher than the cutoff frequency of the $TE_{1,1}$ mode, and the frequency range over which only the dominant

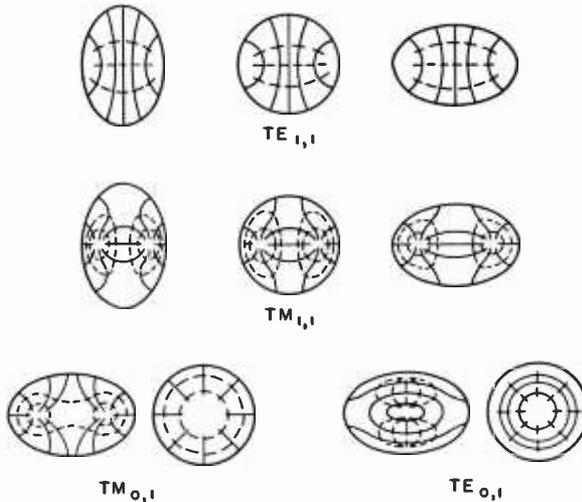


FIG. 7. Modes in elliptical wave guide and the corresponding modes in circular wave guide.

$TE_{1,1}$ mode is above cutoff is correspondingly limited. To operate a wave guide over a wider frequency band in any mode of transmission, the means of excitation must be chosen so as to excite only the desired mode, or else mode dampers must be employed. These mode dampers may consist of grids of wire designed to short-circuit the undesired modes while leaving the desired mode unaffected. Methods of selective mode excitation are discussed later.

5. Elliptical Wave Guides. The inevitable deformations encountered in circular wave guides result in an equivalent ellipticity, and the properties of elliptical wave

guides are, therefore, of practical interest.¹ A few of the possible modes in elliptical wave guide are shown in Fig. 7, along with the corresponding modes in circular wave guide. In general, if a round wave guide is deformed, each mode will split into two modes of orthogonal polarization with different phase velocities and different attenuations. Only those modes with circular symmetry (*e.g.*, $TE_{0,1}$ and $TM_{0,1}$) do not have this instability.

Other cross sections for wave guides have been the subject of some theoretical investigation but have found limited practical application to date. One of the more interesting designs is the H-shaped cross section² shown in Fig. 8. A wave guide of this design has over-all dimensions that are relatively small compared to the cutoff wavelength. Also, the frequency separation between cutoff for the dominant mode of transmission and for the higher modes is comparatively large.

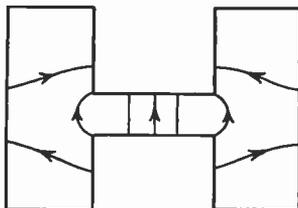


FIG. 8. Wave guide of H-shaped cross section.

6. Characteristic Impedance of Wave Guides. The characteristic impedance of ordinary two-wire transmission lines may be defined as the input impedance to a line of infinite length and may be expressed as the ratio of voltage E between conductors to the current I flowing in each conductor ($Z_0 = E/I$) for a transmission line free from standing waves. It is not possible to extend this definition to wave guides, because

there are no separated conductors between which a unique value of voltage exists. In an ordinary transmission line, it is possible also to define characteristic impedance in terms of the net power flow W into a matched line ($Z_0 = W/I^2$ or $Z_0 = E^2/W$). The net power flow into an infinite wave guide may be calculated when the field strengths at a cross section are given. If a certain logical value of voltage or current is then chosen as reference, it is possible to define a characteristic impedance of the guide in terms of the reference voltage or current and the power input. The actual numerical value of the impedance will depend upon precisely what voltage or current was chosen for reference. The ratio between values of characteristic impedance obtained by different choices of reference voltage or current will be numerically constant, however, and independent of changing frequency or guide dimensions.

For example, with the $TE_{1,0}$ mode in rectangular wave guide, if the maximum voltage between top and bottom guide faces is chosen as the reference voltage, the characteristic impedance of the guide becomes³

$$Z_0 = 377 \frac{\lambda_g 2b}{\lambda a} \quad (8)$$

If the total axial current in the top or bottom face is chosen as the reference current, the characteristic impedance will differ from the above expression only by the numerical factor of $\pi^2/16$.

For many applications, the numerical value of the characteristic impedance is of little importance, as the impedances of the associated elements in the circuit may be normalized with respect to the guide impedance.

7. Power-carrying Capacity of Wave Guides. The power-carrying capacity depends upon the maximum electric field strength that can exist without breakdown. Experimental evidence indicates that at microwave frequencies, this maximum field strength is 30,000 volts per cm under standard sea-level conditions of temperature, pressure, and humidity.

¹ Schelkunoff, S. A., Note on Certain Guided Waves in Slightly Noncircular Tubes, *J. Applied Phys.*, 9 (7), 484, July, 1938. Chu, L. J., Electromagnetic Waves in Elliptic Hollow Pipes of Metal, *J. Applied Phys.*, 9 (9), 583, September, 1938.

² Ramo, S., and J. R. Whinnery, "Fields and Waves in Modern Radio," p. 349, Wiley, 1944.

³ Schelkunoff, S. A., Impedance Concept in Wave Guides, *Quart. Applied Math.*, 2 (1), 1-15, April, 1944.

For rectangular wave guide operating in the $TE_{1,0}$ mode, the maximum power P is related to the maximum field strength E_{max} by

$$\frac{P}{E_{max}^2} = 6.63 \times 10^{-4} ab \left(\frac{\lambda}{\lambda_g}\right) \tag{9}$$

If E_{max} is given in volts per centimeter, the inside dimensions of the guide, a and b , should be given in centimeters for the power to be in watts. The free-space wavelength is λ , the guide wavelength λ_g . The maximum field intensity occurs parallel to the narrower dimension of the guide, midway between the side walls, and is independent of the distance from the wide faces of the guide.

For circular wave guides operating in the dominant $TE_{1,1}$ mode the relation between maximum power and maximum field strength is

$$\frac{P}{E_{max}^2} = 1.99 \times 10^{-3} a^2 \left(\frac{\lambda}{\lambda_g}\right) \tag{10}$$

- where a = radius of guide, cm
- λ = free-space wavelength
- λ_g = wave-guide wavelength

Maximum field strength is at the center of the guide.

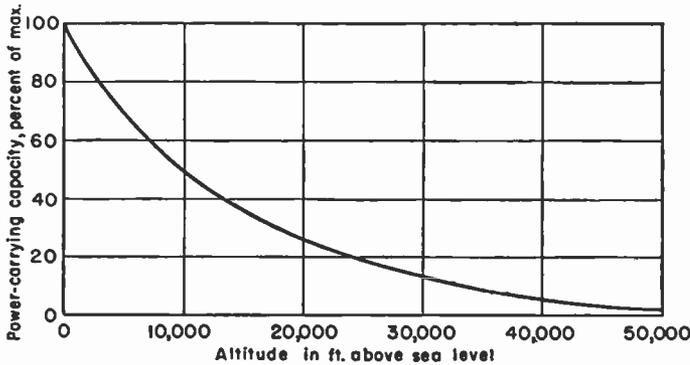


FIG. 9. Decrease in power-carrying capacity of wave guide with increasing altitude (decreasing pressure).

For the $TM_{0,1}$ mode in circular wave guide, there are two cases:

Case 1— $a/\lambda < 0.761$:

$$\frac{P}{E_{max}^2} = 7.69 \times 10^{-3} \frac{a^4}{\lambda^2} \left(\frac{\lambda}{\lambda_g}\right) \tag{11}$$

The maximum field intensity is at the center of the guide.

Case 2— $a/\lambda > 0.761$:

$$\frac{P}{E_{max}^2} = 3.33 \times 10^{-3} a^2 \left(\frac{\lambda_g}{\lambda}\right) \tag{12}$$

The maximum field intensity is at a radius of $0.765a$ and is independent of angle. All dimensions in the above formulas should be given in centimeters for the power to be in watts.

The power-carrying capacity of a wave guide will diminish with altitude approximately as shown in Fig. 9.

8. Attenuation in Wave Guides. At frequencies below cutoff, the attenuation in any wave guide is very rapid, even for perfectly conducting walls, as the guide is unable to support traveling waves. At frequencies higher than cutoff, the guide can

support traveling waves, but they are attenuated because of losses in the conducting walls and in the dielectric that fills the guide. For air-filled guides, the dielectric losses are usually negligible, but if dielectrics other than air are used, the dielectric losses are often greater than the conductor losses.

9. Conductor Losses. The conductor losses in a wave guide are governed in part by skin depth considerations, as the current that flows in the guide walls is concentrated near the inner surface. The attenuation will vary as the square root of the resistivity of the material which forms the conducting walls.

As the frequency increases beyond the cutoff frequency, the attenuation drops rapidly from a very high value at cutoff to a broad minimum, and then increases slowly with increasing frequency. This is the behavior for all modes in all wave guides except the family of circular electric modes (*e.g.*, $TE_{0,1}$) in circular wave guide. These modes have anomalous characteristics that are discussed in a succeeding paragraph.

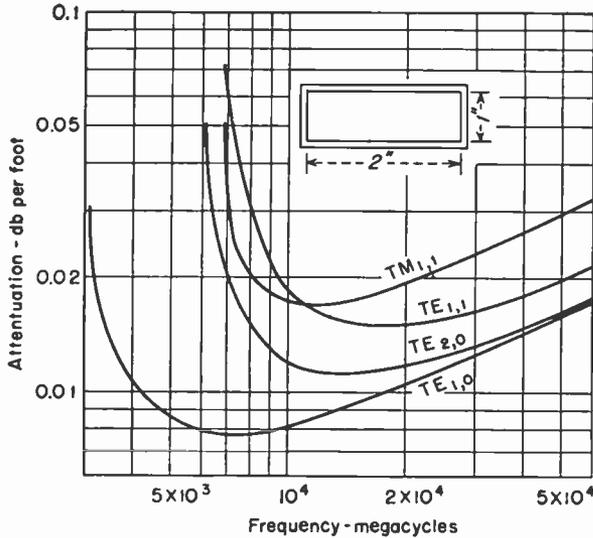


FIG. 10. Attenuation in typical rectangular, air-filled copper wave guide.

The formulas that are given below for wave-guide attenuation indicate that the attenuation approaches infinity as the frequency approaches cutoff. However, the approximations under which these equations were developed are not valid in the region immediately adjacent to cutoff, for there is actually a smooth transition between the approximate results that are given below for the above-cutoff and below-cutoff regions.

10. Attenuation in Rectangular Wave Guides. Formulas for the attenuation in rectangular, copper, air-filled wave guide are given in Table 3. The inner dimensions of the guide, a and b , are in inches. The larger dimension is a . If some metal other than copper is used as a conductor, the attenuation given by Table 3 should be multiplied by the square root of the resistivity of the metal to that of copper.

The variation of attenuation with frequency in a typical rectangular wave guide is given in Fig. 10 for some of the lower modes of propagation. In Fig. 11, the variation of attenuation with frequency is given for the dominant $TE_{1,0}$ mode for a number of widely used guide sizes over the frequency range for which the guide is recommended.

11. Circular Wave Guides. Attenuation formulas for various modes in circular, copper wave guide are given in Table 4. The guide radius a should be given in inches. The variation of attenuation with frequency for each of the above modes in a copper

Table 3. Attenuation in Rectangular, Copper, Air-filled Wave Guides

b

a

f = frequency
 f_c = cutoff frequency
 a and b in inches

Mode	α , Db/Ft
$TE_{1,1}$	$\frac{0.01107}{a^{3/2}} \left[\frac{\frac{1}{2} \frac{a}{b} \left(\frac{f}{f_c}\right)^{3/2} + \left(\frac{f}{f_c}\right)^{-1/2}}{\sqrt{\left(\frac{f}{f_c}\right)^2 - 1}} \right]$
$TE_{2,0}$	$\frac{0.01565}{a^{3/2}} \left[\frac{\frac{1}{2} \frac{a}{b} \left(\frac{f}{f_c}\right)^{3/2} + \left(\frac{f}{f_c}\right)^{-1/2}}{\sqrt{\left(\frac{f}{f_c}\right)^2 - 1}} \right]$
$TE_{1,1}$	$\frac{0.01107}{a^{3/2}} \left\{ \frac{\frac{a}{b} \left(1 + \frac{a}{b}\right) \left(\frac{f}{f_c}\right)^{3/2} + \left[1 + \left(\frac{a}{b}\right)^2\right] \left(\frac{f}{f_c}\right)^{-1/2}}{\left[1 + \left(\frac{a}{b}\right)^2\right]^{3/4} \sqrt{\left(\frac{f}{f_c}\right)^2 - 1}} \right\}$
$TM_{1,1}$	$\frac{0.01107}{a^{3/2}} \left\{ \frac{\left[1 + \left(\frac{b}{a}\right)^2\right] \left(\frac{f}{f_c}\right)^{3/2}}{\left[1 + \left(\frac{b}{a}\right)^2\right]^{3/4} \sqrt{\left(\frac{f}{f_c}\right)^2 - 1}} \right\}$

wave guide of 2 in. inside diameter is given in Fig. 12. Note that the attenuation of the $TE_{0,1}$ mode decreases without limit with increasing frequency.¹ Experimental verification of this is lacking, and this anomalous attenuation characteristic is lost if the guide is elliptical. Nevertheless, this mode offers the lowest attenuation of any of the lower wave-guide modes, provided troublesome effects from other modes which are above cutoff can be eliminated. The changes in attenuation caused by small amounts of ellipticity are small in magnitude unless the frequency is far above cutoff.

12. Comparison of Theoretical and Measured Attenuations. The values of attenuation given in the preceding paragraphs are all derived from theoretical considerations. Because the current in a wave guide is concentrated at the inner surface of the guide, the condition of this surface affects the attenuation, which is usually higher than predicted theoretically. The difference will generally be small at the low end of the microwave spectrum, but at higher frequencies surface roughness will increase the attenuation by increasing the path length which the surface current must travel. For example, at 25,000 Mc, even a well-machined surface will frequently have only two-thirds the theoretical conductivity of an ideal surface. Drawn brass tubing usually approaches rather closely the theoretical values at most microwave frequencies.

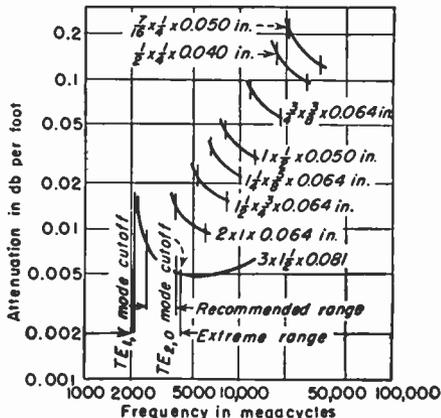
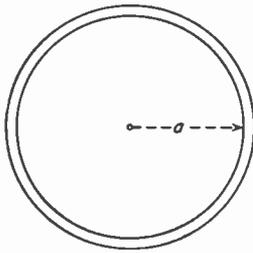


Fig. 11. Theoretical attenuation of copper wave guide over the recommended operating frequency range for the dominant mode. Dimensions are outside measurements and wall thickness.

¹ Kemp, J., Anomalous Attenuation in Waveguides, *Wireless Engr.*, 23 (275), 211, August, 1946.

Table 4. Attenuation in Circular, Copper Wave Guides



f = frequency
 f_c = cutoff frequency
 Radius a in inches

Mode	$\alpha, \text{Db/Ft}$
$TE_{1,1}$ (dominant)	$\frac{0.00423}{a^{3/2}} \left[\frac{\left(\frac{f}{f_c}\right)^{-1/2} + \frac{1}{2.38} \left(\frac{f}{f_c}\right)^{3/2}}{\sqrt{\left(\frac{f}{f_c}\right)^2 - 1}} \right]$
$TM_{0,1}$ (circular magnetic)	$\frac{0.00485}{a^{3/2}} \left[\frac{\left(\frac{f}{f_c}\right)^{3/2}}{\sqrt{\left(\frac{f}{f_c}\right)^2 - 1}} \right]$
$TE_{0,1}$ (circular electric)	$\frac{0.00611}{a^{3/2}} \left[\frac{\left(\frac{f}{f_c}\right)^{-1/2}}{\sqrt{\left(\frac{f}{f_c}\right)^2 - 1}} \right]$

The skin depth in most wave guides is so small that if the surface of the guide is

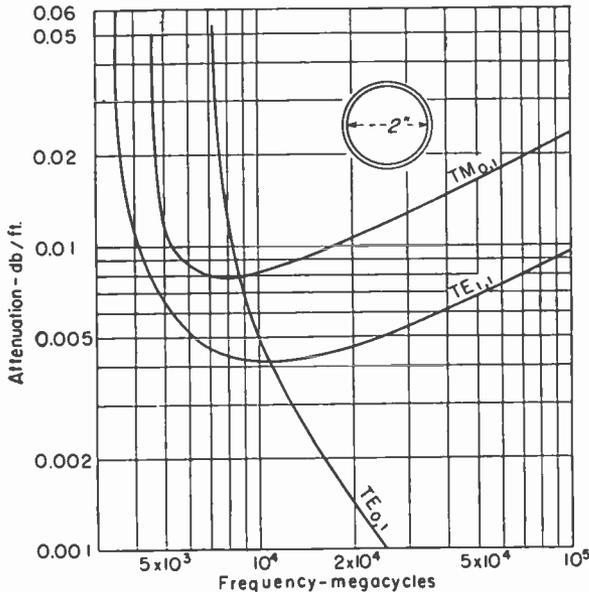


FIG. 12. Attenuation in typical circular, air-filled copper wave guide.

plated to a thickness of a thousandth of an inch or less, essentially all of the surface current will flow in the plated surface metal. But because of the porosity of the plated surfaces, the attenuation will be somewhat higher than expected from the l-f resistivity of the plated metal. The condition of the surface before plating and the plating technique will have considerable effect upon the attenuation. For these reasons, attenuation in plated wave guides varies considerably from sample to sample, and at 10,000 Mc will frequently be half again as great as the theoretical value.

If the inside of a wave guide is coated with a thin protective coating of high-resistance material,

such as protective lacquer, the current will flow under this layer in the base metal, and the attenuation will not be greatly affected.

13. Attenuation in Wave Guides below Cutoff. At frequencies below cutoff, wave guides are unable to support traveling waves which carry energy down the guide, and if modes are excited by some launching mechanism, the field intensity will diminish very rapidly with increasing distance from the point of excitation.¹ The input impedance of a wave guide below cutoff is a pure reactance.

For any one mode in the wave guide, the field strength as a function of the distance from the point of excitation will vary as

$$E = E_0 e^{-\alpha z} \tag{13}$$

where E_0 is the initial amplitude at the point of excitation. The attenuation is given by

$$\alpha = 8.69 \sqrt{\left(\frac{2\pi}{\lambda_c}\right)^2 - \epsilon \left(\frac{2\pi}{\lambda}\right)^2} \quad \text{db/unit length} \tag{14}$$

where λ_c = cutoff wavelength of air-filled guide

λ = wavelength in free space at frequency of excitation

ϵ = dielectric constant of dielectric medium inside wave guide

At frequencies much lower than the cutoff frequency, the attenuation is very nearly independent of frequency, approaching a limiting value at low frequencies of

$$\alpha = \frac{54.6}{\lambda_c} \quad \text{db/unit length} \tag{15}$$

For the two modes of greatest interest in circular wave guide, the modes that are commonly used in wave guide below-cutoff attenuators, the attenuations are as follows:

For the $TE_{1,1}$ mode,

$$\alpha = 8.69 \sqrt{\left(\frac{1.841}{a}\right)^2 - \left(\frac{2\pi}{\lambda}\right)^2} \quad \text{db/unit length} \tag{16}$$

For the $TM_{0,1}$ mode,

$$\alpha = 8.69 \sqrt{\left(\frac{2.405}{a}\right)^2 - \left(\frac{2\pi}{\lambda}\right)^2} \quad \text{db/unit length} \tag{17}$$

where a is the guide radius and λ the free-space wavelength.

14. Wave Guides Filled with Dielectric Material. If a wave guide is partially or completely filled with some dielectric material, the characteristics of the guide will be modified. In dealing with these modifications, it is convenient to regard the dielectric constant ϵ of the material as complex and of the form

$$\epsilon = \epsilon' - j\epsilon'' \tag{18}$$

The loss tangent of the dielectric is then defined by

$$\tan \delta = \frac{\epsilon''}{\epsilon'} \tag{19}$$

For all but high-loss dielectrics, the loss tangent is very nearly equal to the power factor, and the two terms are frequently used interchangeably.

For all modes in wave guides, the cutoff wavelength of a dielectric-filled guide λ_c' is related to the cutoff wavelength of the same guide when air-filled λ_c by

$$\frac{\lambda_c'}{\lambda_c} = \sqrt{\epsilon} \tag{20}$$

¹Linder, E. G., Attenuation of Electromagnetic Fields in Pipes Smaller Than the Critical Size, *Proc. IRE*, 30 (9), 412, September, 1942.

The wave-guide wavelength in dielectric-filled guide for all modes is given by

$$\lambda_g = \frac{\lambda}{\sqrt{\epsilon' - (\lambda/\lambda_c)^2}} \tag{21}$$

where λ_c is the cutoff wavelength in the same guide when air-filled.

The standing-wave ratio associated with the reflection from an air-dielectric interface normal to the axis of the wave guide is, for *TE* modes

$$SWR = \sqrt{\frac{1 - (\lambda/\lambda_c)^2}{\epsilon' - (\lambda/\lambda_c)^2}} \tag{22}$$

and for *TM* modes,

$$SWR = \frac{1}{\epsilon'} \sqrt{\frac{\epsilon' - (\lambda/\lambda_c)^2}{1 - (\lambda/\lambda_c)^2}} \tag{23}$$

This reflection can be minimized by tapering the interface in the axial direction.

The attenuation in a below-cutoff wave guide filled with dielectric is given in Eq. (14).

In a wave guide above cutoff filled with dielectric, the attenuation resulting from dielectric losses is given by

$$\alpha = 27.2 \frac{\epsilon''}{\lambda} \frac{1}{\sqrt{\epsilon' - (\lambda/\lambda_c)^2}} \text{ db/unit length} \tag{24}$$

The conductor losses are also affected by the dielectric. To find the conductor losses in a dielectric-filled wave guide, it is necessary to modify the expressions in Tables 3 and 4 as follows: (1) In all expressions, replace f_c , the cutoff frequency of the air-filled guide, by $f_c/\sqrt{\epsilon}$, the cutoff frequency of the dielectric-filled guide; (2) multiply the resulting expression by $\epsilon^{1/4}$.

The total attenuation in a dielectric-filled wave guide is the sum of the attenuation resulting from dielectric losses and that resulting from conductor losses.

15. Wave Guides Partially Filled with Dielectric.¹ If a wave guide is partially filled with a dielectric, its properties will, as expected, be intermediate between airfilled and dielectric-filled guides. The fields inside the guide tend to be pulled inside the medium with the higher dielectric constant. The distribution of the fields for a wave guide partially filled with dielectric is shown in Fig. 13, and it is apparent how the point of maximum electric field strength,

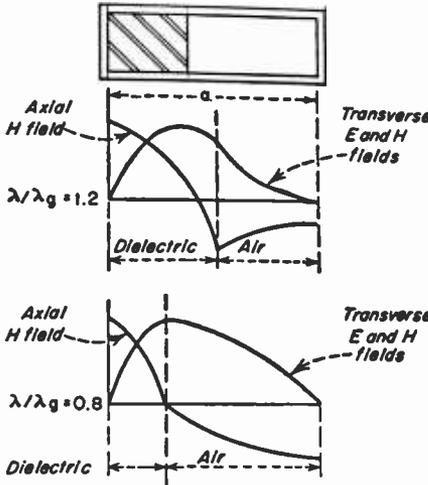


FIG. 13. Field intensities in a rectangular wave guide partially filled with a dielectric operating in the *TE*_{1,0} mode.

normally in the center of the guide, has been pulled toward the dielectric.

Placing the dielectric inside the guide will increase the cutoff wavelength and reduce the wave-guide wavelength, but the magnitude of the effect will depend upon the loca-

¹ Pincherle, L., *Electromagnetic Waves in Metal Tubes Filled Longitudinally with Two Dielectrics*, *Phys. Rev.*, 66 (5, 6), Sept. 1 and 15, 1944.

tion of the dielectric. The effects will be small, for example, if the dielectric is placed in a region of a weak electric field, as shown in Fig. 14, but will be much greater if the dielectric is placed in the center of the guide, in a region of strong field.

When the dielectric is tapered to minimize the reflection from a transition between dielectric-filled and air-filled guides, it is advisable to start and end the taper in the regions of the weak field. For example, with the $TE_{1,0}$ mode in rectangular wave guides, the best results are found with a taper constructed as shown in Fig. 15.

16. Dielectric Wave Guides. It is possible for a dielectric cylinder immersed in a medium of lower dielectric constant to act as a wave guide in the absence of any conducting surface.¹ Such a wave guide without metal walls has many properties similar to more conventional wave guides; energy is carried in TE or TM modes, and the guide acts as a high-pass filter, with the cutoff frequency different for the different modes. But the electromagnetic fields are not confined to the inside of the dielectric cylinder, and a considerable fraction of the total energy

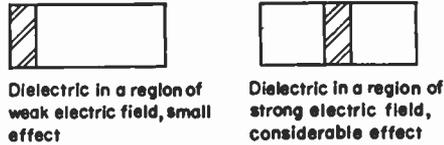


FIG. 14. A dielectric placed in a region of strong electric field will have a greater effect upon the properties of the wave guide than if it is placed in region of weak field.

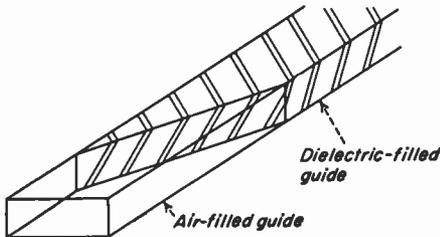


FIG. 15. Impedance discontinuity at the junction of air-filled wave guide and dielectric-filled wave guide can be minimized by tapering the dielectric.

17. Obstacles and Discontinuities in Wave Guides. A metallic obstacle placed in a wave guide will in general affect the transmission of an electromagnetic wave through the guide. The obstacle will cause a partial reflection of the incident wave and, in addition, will excite other of the wave-guide transmission modes. If these are also above cutoff, part of the energy in the incident wave will be carried away from the discontinuity by these other modes of transmission. If the modes excited by the discontinuity are below cutoff, they will diminish rapidly in strength with increasing distance from the discontinuity and will draw no real power. They do contribute a reactive effect at the discontinuity, however.

The effect of an obstacle in a wave guide is equivalent to that of a discontinuity in a conventional transmission line, and it is convenient for engineering work to set up an equivalent transmission line and lumped constant circuit for the wave-guide structure. This equivalent circuit will in general be a function of frequency.

If the obstacle has axial dimensions that are very small, *e.g.*, a thin metallic window that partially blocks the guide, the equivalent circuit will be a simple susceptance shunting the guide that is located precisely at the point of discontinuity.² If the obstacle is of appreciable thickness, a more complex equivalent circuit in the form of a

¹ For example, see Ramo and Whinnery, *loc. cit.*

² Whinnery, J. R., and H. W. Jamieson, Equivalent Circuits for Discontinuities in Transmission Lines, *Proc. IRE*, 32 (2), 98, February, 1944. Also Schelkunoff, S. A., "Electromagnetic Waves," p. 490, Van Nostrand, 1943.

tee or pi network is usually required, the precise circuit depending upon what planes in the wave guide are chosen for reference.

If two obstacles are placed close enough together for interaction of the below-cutoff higher order modes, the equivalent circuits will be correspondingly modified. Inter-

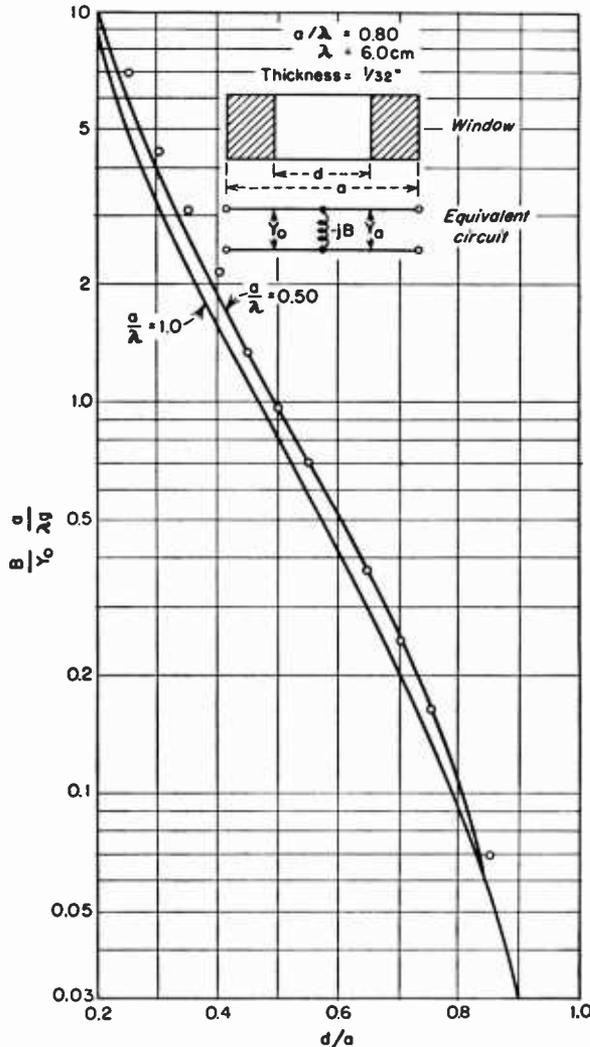


FIG. 16. Equivalent susceptance of a thin symmetrical inductive window in a rectangular wave guide. Theoretical curves, experimental points.

action effects are usually small if the spacing is greater than a quarter-wave-guide wavelength.

The equivalent circuits given in this chapter are in general applicable when the wave guide is above cutoff for the dominant mode only.

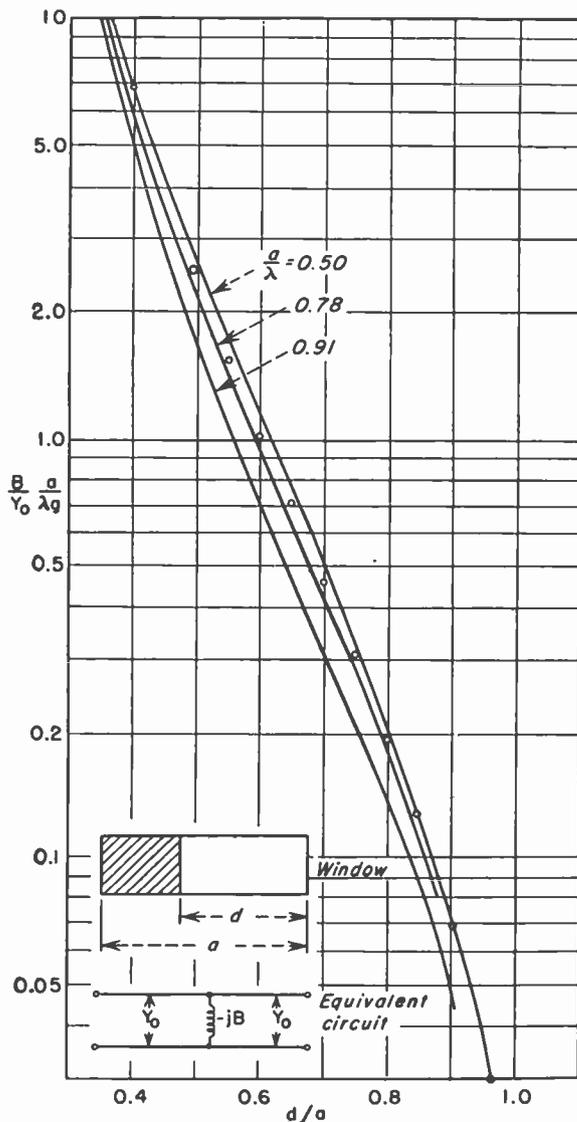


FIG. 17. Equivalent susceptance of thin, asymmetrical, inductive window in rectangular wave guide.

18. **Obstacles in Rectangular Wave Guide. Thin Metallic Windows.**¹ A thin metallic window that partially blocks the guide has an equivalent circuit consisting of a susceptance shunting the guide at the location of the window. If the window is of the type shown in Fig. 16, the equivalent circuit is an inductive susceptance, whose

¹ Miles, J. W., The Equivalent Circuit for a Plane Discontinuity in a Cylindrical Wave Guide, *Proc. IRE*, 34 (10), 728, October, 1946.

magnitude is given in Fig. 16 as a function of the window dimensions. The curves are for infinitely thin windows, the experimental points for windows of $\frac{1}{32}$ in. thickness at a wavelength of 6.0 cm. The effect of finite thickness is to increase the susceptance of the window, and series reactance terms in the equivalent circuit are also necessary if the thickness is considerable.

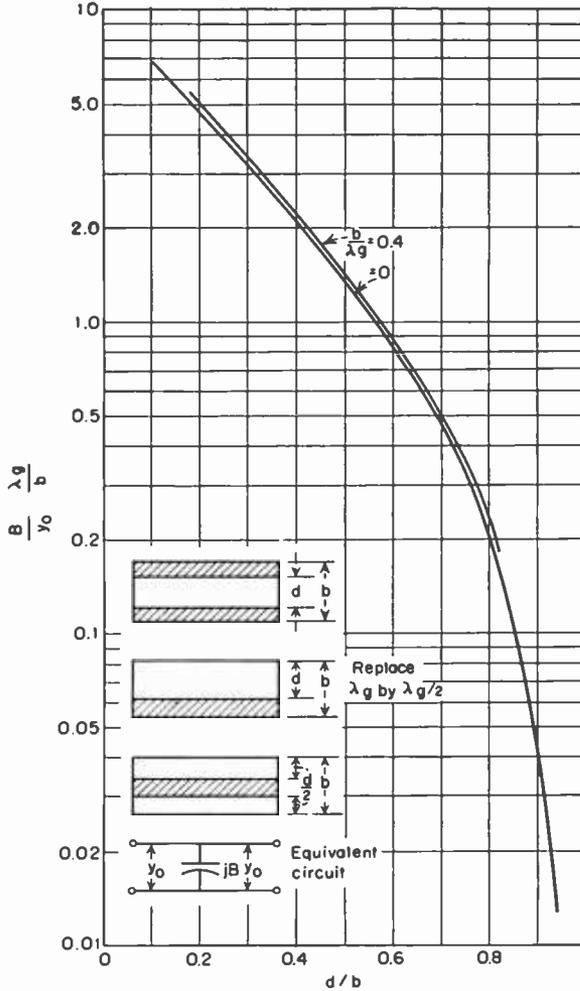


FIG. 18. Equivalent susceptance of a thin capacitive window in rectangular wave guide. Theoretical curves for infinitely thin window.

If the opening in the diaphragm is of the type shown in Fig. 17, the equivalent circuit is again a shunt inductive susceptance. The susceptance as a function of the dimensions is given in Fig. 17, along with some experimental results, taken with a window $\frac{1}{32}$ in. thick at a wavelength of 6.0 cm.

If the window is of the type shown in Fig. 18, the equivalent circuit is a capacitive

susceptance shunted across the guide. The theoretical susceptance for an infinitely thin window is given in Fig. 18. With a window of this type, it is very important for accurate results to take into account the finite thickness of the window. If the window thickness is finite but much less than a wavelength, its susceptance will be given by

$$\frac{B}{Y_0} = \frac{B_0}{Y_0} + \frac{2\pi t}{\lambda_0} \left(\frac{b}{d} - \frac{d}{b} \right) \quad (25)$$

where B_0/Y_0 = thin window susceptance taken from Fig. 18

λ_0 = wave-guide wavelength
 t = window thickness

A window of the type shown in Fig. 19 exhibits susceptance vs. frequency characteristics similar to those of a parallel resonant circuit shunting the wave guide. The resonant wavelength, at which the window introduces no reflection, may be determined approximately from the expression

$$\frac{a}{b} \sqrt{1 - \left(\frac{\lambda}{2a} \right)^2} = \frac{a'}{b'} \sqrt{1 - \left(\frac{\lambda}{2a'} \right)^2} \quad (26)$$

where a and b = inside guide dimensions

a' and b' = window dimensions, as shown in Fig. 19

λ = free-space wavelength

The slope at resonance of the susceptance vs. frequency curve decreases as the height of the window opening increases.

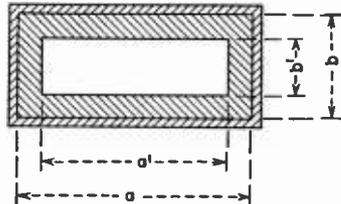


FIG. 19. Resonant window in a thin metallic diaphragm.

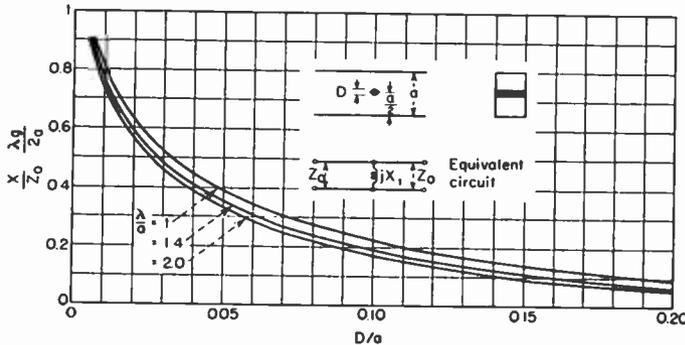


FIG. 20. Equivalent susceptance of round metallic post in a rectangular wave guide.

Metallic Posts. A round metallic post symmetrically located in a rectangular wave guide, as shown in Fig. 20, has an equivalent circuit consisting of an inductive susceptance shunted across the guide at the location of the post. The magnitude of this susceptance is given as a function of the post diameter.

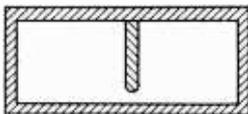


FIG. 21. Resonant metallic post in rectangular wave guide.

If the post extends only part way across the guide, as in Fig. 21, the equivalent circuit is a capacitive susceptance, whose magnitude depends upon the length and diameter of the post. At any one frequency, as the length of the post is increased, a position is found at which the susceptance becomes infinite in magnitude. As the frequency is

varied about this value, the variation of susceptance vs. frequency is similar to that of a series-resonant circuit shunting the wave guide. The resonant wavelength is

found when the post length is approximately one-quarter of a free-space wavelength, slightly less for posts of larger diameter.

Resonant Rings. A rectangular metal ring placed in a rectangular wave guide perpendicular to the axis also exhibits characteristics similar to a series-resonant circuit shunting the guide. The resonant wavelength depends upon the mean perimeter of the ring and is found when the perimeter is somewhat greater than a full wavelength.

19. Obstacles in Circular Wave Guides. For a wave guide which is above cutoff for the $TE_{1,1}$ mode only, a thin diaphragm with a centered, circular hole behaves as a shunting inductive susceptance. The magnitude of this susceptance as a function of the hole diameter is given in Fig. 22. A thin metal disk centrally located in the wave guide behaves as a capacitive susceptance shunted across the guide.

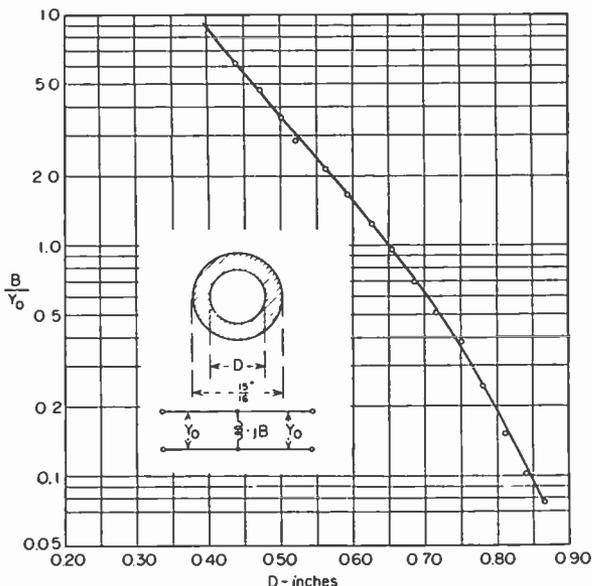


FIG. 22. Measured equivalent susceptance of a circular hole in a thin diaphragm in a circular wave guide operating in the $TE_{1,1}$ mode. The operating wavelength was 3.2 cm; window thickness, $\frac{1}{32}$ in.

A variety of apertures in thin diaphragms act as parallel resonant circuits shunting the guide, and for each of these apertures it is usually possible to find a similar obstacle which acts as a series-resonant circuit shunting the guide. Some examples of these circuits are shown in Fig. 23.

20. Miscellaneous Wave-guide Structures.¹ Bends. The direction of transmission in a wave guide may be changed by bending the guide. The bent section has characteristics slightly different from the straight guide, and consequently there will be an impedance discontinuity at the junction of the bent and straight sections. A number of modes will be excited at the discontinuity and may carry away appreciable power if the guide is above cutoff for these other modes.

If the guide is operating with the dominant mode and if all other modes are below cutoff, a bend is usually a very satisfactory way of changing direction of transmission. For rectangular guides, there are two types of bend, shown in Fig. 24, known as E-plane and H-plane bends, in the plane of the electric and magnetic fields, respectively. If the inner radius of the bend is a wave-guide wavelength or greater, the reflection from

¹ Moreno, T., Wave Guide Transmission Systems, *Electronics*, 19 (6), 136, June, 1946.

the bend will be small ($SWR < 1.05$) over the normal operating range of the wave guide. The radius may be much smaller and still yield satisfactory results if the bend is carefully fabricated. With a sharp bend, the cross section of the guide should be undistorted for best operating results.

A bend in a circular wave guide operating in the $TE_{1,1}$ mode, if not in the plane of polarization or normal to that plane, will split the transmitted wave into two components with different phase velocities and cause elliptical polarization of the transmitted wave. If the plane of the bend is in one of the planes of mode symmetry, the radius of the bend can be considerably less than a wave-guide wavelength, and the bend will give satisfactory results if the cross section is not distorted.

Corners. An abrupt corner is in general not a satisfactory means of changing the direction of transmission in a wave guide, as an appreciable reflection will be set up for all but very small changes in direction. If the desired angular change is accomplished with the use of two simple corners, spaced apart approximately a quarter-wave-guide wavelength, the reflections

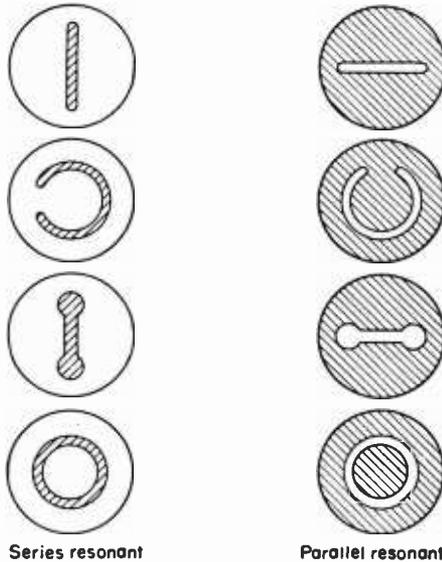
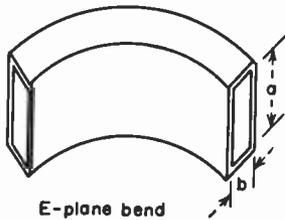
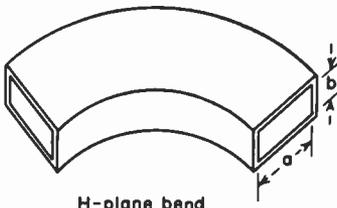


FIG. 23. Wave-guide structures equivalent to resonant circuits shunting the wave guide. Circular wave guide operating in the $TE_{1,1}$ mode with electric field vertically polarized.



E-plane bend



H-plane bend

FIG. 24. Bends in rectangular wave guide.

from the two corners will tend to cancel each other. For E-plane corners, shown in Fig. 25, the optimum spacing is very close to a quarter-wave-guide wavelength at all wavelengths, but for H-plane corners, the optimum spacing is slightly different and is plotted as a function of λ/a in Fig. 25 for 90 deg total change in direction.

The band width over which these corners will introduce a standing-wave ratio less than 1.05 is between 6 and 20 per cent for a 90-deg bend, being larger for smaller values of λ/a . Circular bends are preferred for larger angles and greater band widths.

The reflection from a corner may also be eliminated by modifying the corner as shown in Fig. 26, and the band width will be comparable to that obtained with the double corner. The required dimensions for E- and H-plane bends are given in Fig. 26 as a function of the angle. Manufacturing tolerances must be held rather closely on these corners for optimum performance.

Twists. The direction of polarization in a rectangular wave-guide transmission system may be changed by twisting the guide about its axis (Fig. 27). If carefully fabricated, these twists will introduce no appreciable reflection into a matched transmission system. A 90-deg

will introduce no appreciable reflection into a matched transmission system. A 90-deg

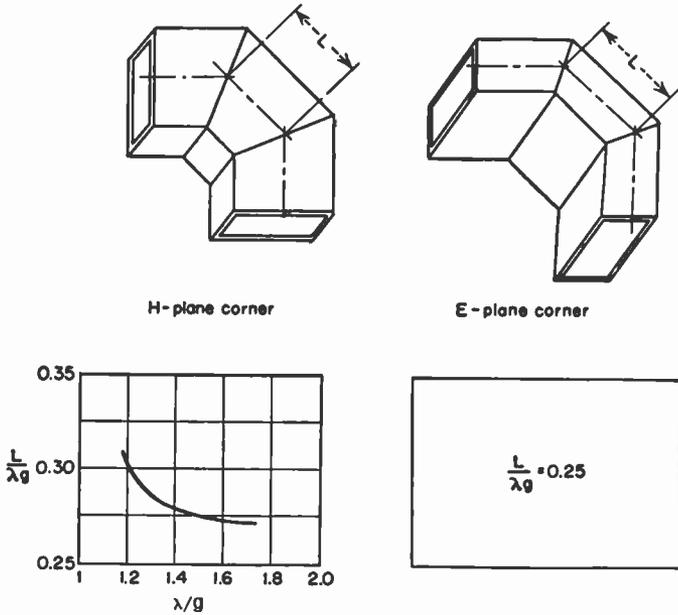


FIG. 25. Corners, 90 deg, designed for minimum reflection. Optimum dimensions for H-plane corners in graph; for E corners, in box.

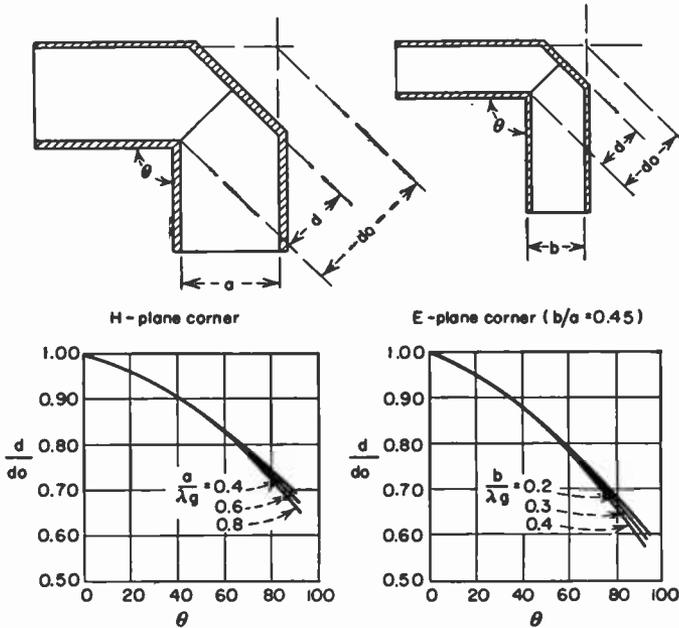


FIG. 26. Minimum reflection corners with dimensions as function of angle.

twist will generally introduce a standing-wave ratio below 1.1 if more than 2λ , in length, and shorter twists are equally satisfactory if carefully made.

21. Wave-guide Connectors. Two sections of a wave guide may be connected by machining the two ends of the guide off square and clamping them together. This is a very satisfactory connection if the guides are carefully aligned and tightly clamped, and the butt joint will be electrically smooth and free from loss. The efficiency of the connection depends critically upon the accuracy of the joint, however. If the two guides are slightly misaligned or not tightly clamped, or if the ends are not finished off square and smooth, the loss and reflection may be relatively large. Losses of up to 1.0 db have been measured on butt connections that appeared satisfactory to a superficial examination. So while butt joints may be very satisfactory and are even preferred when accurate measurements are being made, they must be assembled with care and accuracy to give satisfactory performance.

The assembly of wave-guide components is facilitated with the use of choke couplings, as shown in Fig. 28. These offer advantages in flexibility and reliability over simple butt joints because their performance does not depend critically upon the tightness of the connection. The L-shaped cavity between the choke and flange is effectively a half-wave, shorted transmission line, which is in series with the guide and offers a minimum discontinuity. The circular slot in the choke is $\lambda/4$ deep, and the actual contact between choke and flange is in series with the high-input impedance of this resonant slot, and is at a point of minimum current. For this reason, a poor contact between choke and flange does not impair the efficiency of the connection. The distance between the quarter-wave slot and the inside of the wave guide must be determined experimentally, but is effectively $\lambda/4$. For maximum band width, the slot width should be several times greater than the spacing between choke and flange faces. A properly designed choke will introduce a standing-wave ratio less than 1.05 over most of the usable range of a rectangular wave guide.

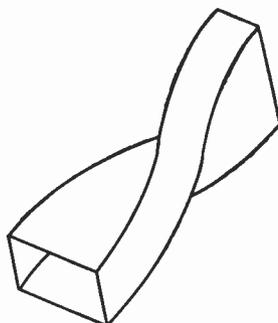
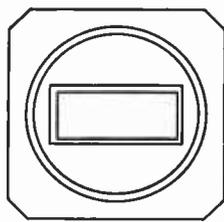


FIG. 27. Twist in rectangular guide to change direction of polarisation.



Face of choke

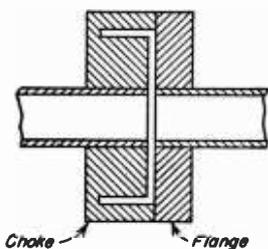


FIG. 28. Choke-flange connector for rectangular wave guide.

ance of $\frac{1}{16}\lambda$ is allowed between choke and flange, a sideways displacement of up to $\frac{1}{16}\lambda$ will not raise the loss above 0.3 db, or the standing-wave ratio above 1.3, for frequencies within a few per cent of the design frequency. A pair of chokes may be used instead of the choke-flange combination for a wobble joint and may have superior performance. With paired chokes there are often resonances encountered near the design wavelength, which manifest themselves by sharp increases in the loss and standing-wave ratio.

22. Launching of Waves in Wave Guides. Energy traveling down a conventional transmission line may be transferred to a wave guide by terminating the transmission line in an antenna structure, which, instead of radiating energy into free space, is

It is not necessary that the outer faces of the choke and flange be in contact for good performance. The connectors may be used as nonrigid couplings, such as are required when a shock-mounted chassis is connected to a rigidly mounted line. If a clear-

mounted inside the wave guide. The transition from conventional transmission line to wave guide may be made electrically smooth and reflectionless by designing the antenna so that it presents a matched load to the transmission line. By the law of reciprocity, a wave traveling from the wave guide to the transmission line will likewise not be reflected, provided the loss in the transition section is small.

The case of greatest practical interest is the launching of a $TE_{1,0}$ mode in rectangular wave guide from a coaxial transmission line. The axis of the coaxial line is generally made normal to the broad face of the rectangular wave guide. As illustrated in Fig.

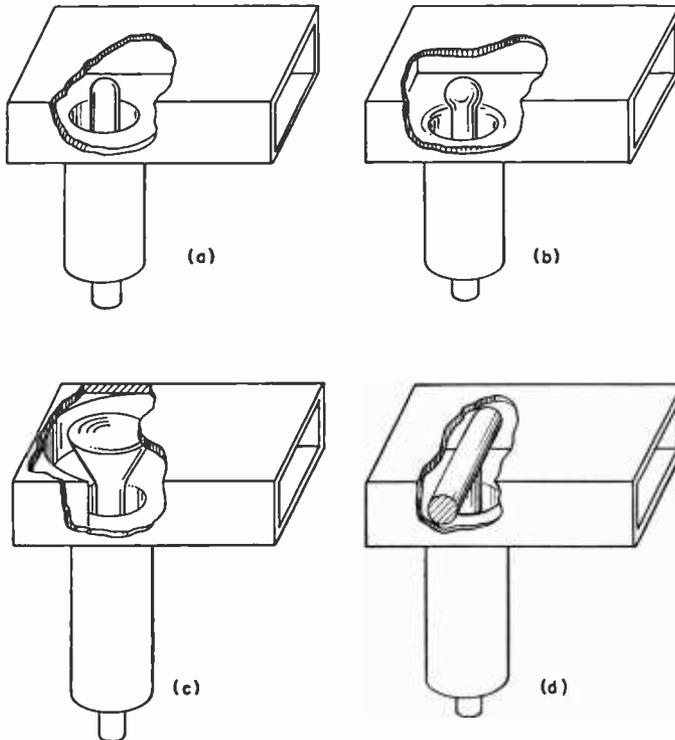


FIG. 29. Broad-band, fixed-tuned coaxial line to wave-guide transformers providing reflectionless transmission.

29a, the outer conductor of the coaxial line is terminated at the wall of the wave guide, but the inner conductor extends into the wave guide, parallel to the electric lines of force, and forms a probe antenna which radiates down the guide.¹

If the wave guide is terminated in both directions in its characteristic impedance, the power input from the coaxial line will divide equally in the two directions. To send the energy in a single direction down the wave guide, one end is shorted with a metal plate. The distance of this plate from the antenna structure and the length of the antenna are chosen to present a matched impedance to the coaxial line. Experimental determination of the optimum dimensions is usually required.

The band width over which the impedance transition between coaxial line and wave guide is relatively smooth and reflectionless can be increased by increasing the diameter of the probe antenna. Further improvements may be realized by modifying

¹ Moreno, *loc. cit.*

the structure as shown in Fig. 29b. The structures shown in Figs. 29c and d are also broad-band transitions between coaxial line and wave guide, and provide the additional advantage of a mechanical support for the center conductor of the coaxial line. Again, the optimum dimensions for these structures must be determined experimentally.

23. Mode Transformers for Wave Guides. It is sometimes desirable in wave-guide transmission systems to change from one mode of transmission in one wave guide to a second mode of transmission in another guide. These transformers may be designed in a variety of ways, of which the following examples are illustrative:

An example of the simplest type of transformer is that required to change from the $TE_{1,0}$ mode in rectangular wave guide to the $TE_{1,1}$ mode in circular wave guide. Because of the similarity of these two modes, it is only necessary to construct a continuous taper in the guide cross-sectional dimensions, as shown in Fig. 30. If the taper is sufficiently long ($2\lambda_0$ is usually sufficient) the required mode transformation will be accomplished with a minimum discontinuity and reflection in the transmission system.

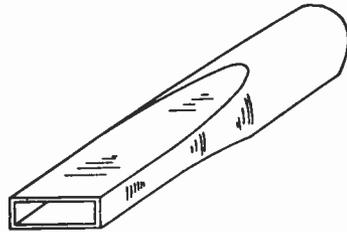


FIG. 30. Tapered transmission between rectangular and circular wave guides, both operating in dominant mode.

A more complex transformer is required for the transition to higher modes in one or both guides, e.g., the transition from $TE_{1,0}$ mode in rectangular guide to $TM_{0,1}$ mode in a circular wave guide. Several transformers have been designed to accomplish this result, some of which are illustrated in Fig. 31. The problem is more difficult because the transition section may excite a large number of modes in the circular wave guide, and the dominant $TE_{1,1}$ mode as well as the desired $TM_{0,1}$ mode will be above cutoff. The transformer must therefore be carefully designed so as to minimize the excitation of the undesired $TE_{1,1}$ mode, whether by reasons of symmetry or by other

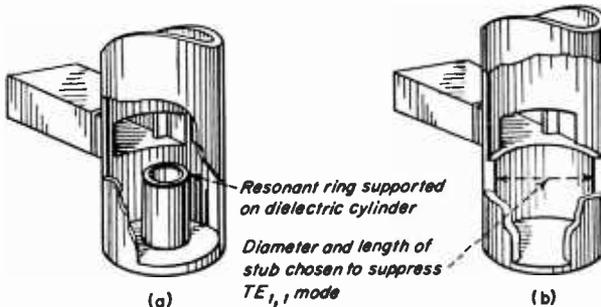


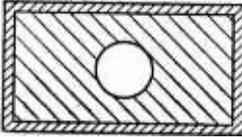
FIG. 31. Transformers from $TE_{1,0}$ mode in rectangular to $TM_{0,1}$ mode in circular wave guide. In addition to providing smooth transition between desired modes, these transformers should not excite the dominant $TE_{1,1}$ mode in the circular wave guide which is above cutoff.

techniques. Mode suppressors are also useful in the design of these transformers. Resonant rings are used for this purpose; when properly designed, they will completely reflect the $TE_{1,1}$ mode and have negligible effect upon the $TM_{0,1}$ mode. One of these resonant rings is shown in the transformer of Fig. 31a, placed in a position to minimize the excitation of the $TE_{1,1}$ mode in the circular wave guide.

In Fig. 31b, the stub section of round wave guide extending below the junction with the rectangular guide is effectively in series with the rectangular and round guides. The length and diameter may be chosen so that the stub is effectively $\lambda/2$ long for the

desired $TM_{0,1}$ mode, but effectively $3\lambda/4$ for the $TE_{1,1}$ mode. The stub therefore offers an infinite impedance in series with the guide for the $TE_{1,1}$ mode, but a zero impedance for the $TM_{0,1}$ mode. In all these transformers, a matching diaphragm may be used to eliminate any residual reflection from the transition.

24. Coupling through Holes between Wave Guides.¹ For a rectangular wave guide operating in the $TE_{1,0}$ mode, the circuit equivalent of a small centered hole in an infinitely thin diaphragm normal to the guide axis is an inductive diaphragm normal to the guide axis is an inductive susceptance shunting the guide (Fig. 32). For dimensions of a and b as shown in Fig. 32, and a hole of diameter d , the susceptance of the hole normalized to the guide admittance is



$$\frac{B}{Y_0} = \frac{3}{2\pi} \frac{ab\lambda_g}{d^3} \tag{27}$$

FIG. 32. Small round hole in thin diaphragm in rectangular guide is equivalent to an inductive susceptance shunting the guide.

The insertion loss of a hole in an infinitely thin diaphragm placed between a matched source and receiver is given by

$$\alpha_1 = 10 \log_{10} \left[\frac{(B/Y_0)^2}{4} - 1 \right] \text{ db} \tag{28}$$

Finite thickness t of the diaphragm results in an additional attenuation, given by

$$\alpha_2 \cong 32.0 \frac{t}{d} \text{ db} \tag{29}$$

The over-all attenuation is then given by $\alpha_t = \alpha_1 + \alpha_2$.

When the hole is used to couple energy between two wave guides whose axes are not coincident, the insertion loss of the coupling hole may in many cases be calculated

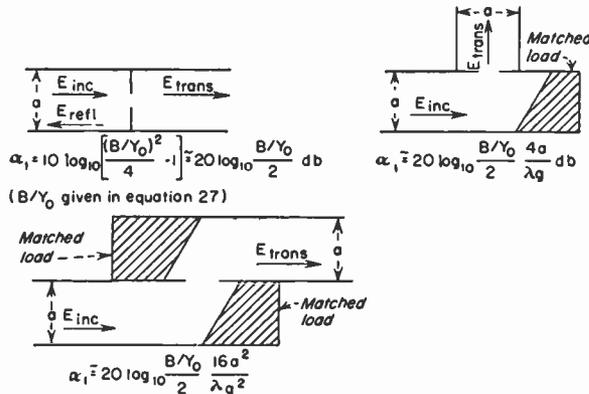


FIG. 33. Attenuation through small hole separating two sections of rectangular guide, under various conditions of operation.

by a simple modification of the term α_1 , of Eq. (28). Examples of arrangements where such simple modifications are possible are shown in Fig. 33. The factor α_2 is not affected by the position of the hole. If the hole is in the broad face of the guides, the problem of coupling is more complex, but a complete solution has been given by Bethe.²

¹ Bethe, H. A., Theory of Diffraction by Small Holes, *Phys. Rev.*, **66** (7, 8), Oct. 1 and 15, 1944.
² Bethe, *loc. cit.*

25. Wave-guide Tees and Junctions.¹ Tee joints in wave guides are of two types, as shown in Fig. 34.

1. *H-plane Tee or Shunt Tee.* The branch guide is taken off the narrow side of the main guide in the plane of the magnetic field. This is called a *shunt tee* because a signal fed in the branch guide will divide at the junction and be in phase in arms 1 and 3 at points equidistant from the junction. The equivalent circuit for two-wire transmission lines is shown in Fig. 34.

2. *E-plane or Series Tee.* The branch guide is taken off the wide side of the main guide in the plane of the electric field. This is called a *series tee* because a signal fed in the branch guide will divide at the junction and be out of phase in arms 1 and 3 at points equidistant from the junction. The equivalent circuit for two-wire transmission lines is shown in Fig. 34.

As would be expected from the equivalent circuits for these tee joints, positions may be found in the branch arms at which short circuits may be placed to give reflectionless transmission past the branch, and other positions may be found at which a short circuit may be placed to give a complete reflection of energy at the branch arm. But because of the distortion and fringing fields at the junction, the positions at which these short circuits must be placed to realize these effects will be somewhat different from the values predicted by conventional transmission-line theory, which neglects these field distortions. For example, with the shunting arm, one would expect that a short circuit should be placed $\lambda/4$ or $3\lambda/4$ away from the junction to give reflectionless transmission past the branch arm. Experimentally, this value is found to vary with ratio of operating wavelength to cutoff wavelength, a typical mid-band value being 0.57λ from short to inside face of the main guide.

Judging by the equivalent circuit, it should be possible to obtain reflectionless transmission around a corner by proper location of a short in the straight-through arm of the wave guide. Experimentally it is found that the reflection may be minimized but not eliminated by proper positioning of the short. This minimum reflection varies with operating wavelength; in typical operation the minimum standing-wave ratio may be approximately 1.3.

To obtain reflectionless transmission around a corner with a tee junction, it is necessary to add some sort of matching structure, such as a diaphragm, the reflection from which is adjusted to cancel the residual reflection from the tee. A typical matching structure of this sort as placed in the branch arm is shown in Fig. 34.

26. Wave-guide Bridge Structures. A variety of wave-guide structures may be built which have properties similar in many respects to l-f bridge circuits. These structures utilize tee joints of the two types discussed in the previous paragraphs.

The best known of these bridge circuits is the *magic tee*, or *hybrid tee*, shown in Fig. 35. Two branch wave guides are connected to a colinear wave guide at the same

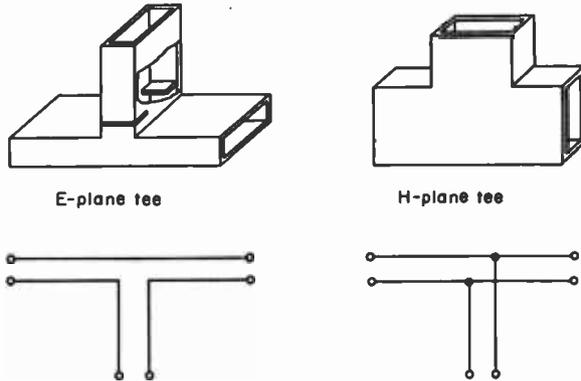


FIG. 34. Two kinds of tee joints in rectangular wave guide and their two-wire transmission line equivalent circuits. The E-plane tee is partially cut away to show an added matching diaphragm.

¹ Allanson, J. T., R. Cooper, T. G. Cowling, The Theory and Experimental Behavior of Right-angled Junctions in Rectangular-section Wave Guides, *J. IEE*, 93 (23), Pt. III, 177.

point, one in an H plane, shunt tee and the other in an E plane, series tee. If a signal is fed in the shunt arm 4, and the two colinear arms 1 and 2 are terminated in their characteristic impedances, the input power will, for reasons of symmetry, divide equally between the two load impedances. Also for reasons of symmetry, there will be no signal delivered into arm 3, as there is no net electric vector developed across

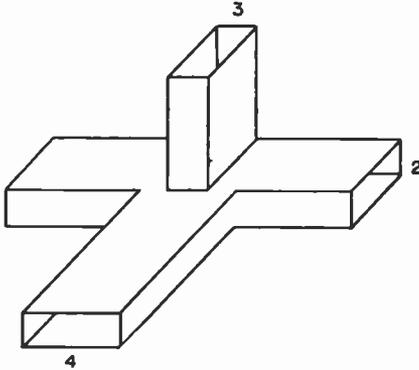


FIG. 35. Magic tee in rectangular guide.

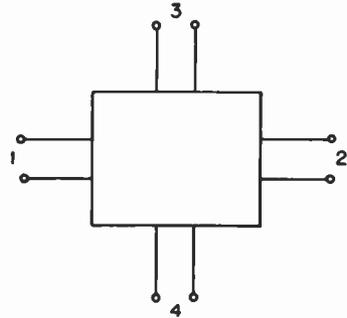


FIG. 36. Symmetrical four-terminal network equivalent of magic tee.

the entrance to this series arm. The input impedance to the shunt arm will not in general match the characteristic impedance of the wave guide; *i.e.*, there will be standing waves in the input arm.

A similar situation will exist if power is fed into the series arm. If the colinear arms 2 and 1 are terminated in their characteristic impedances, the input power will divide equally between these two arms, and no power will be delivered to the shunt branch arm 4. There will be standing waves set up in the input arm, however.

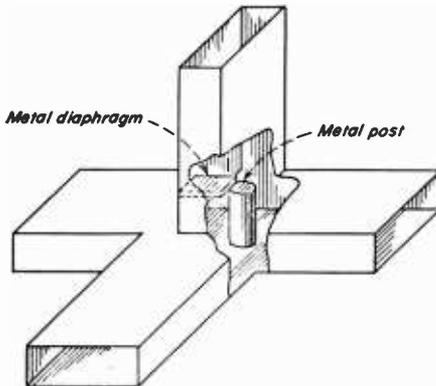


FIG. 37. Cutaway view of magic tee showing diaphragm and post for impedance-matching purposes.

The symmetry of the structure will not be destroyed if the two matched loads on the colinear arms are replaced by any two impedances that are equal in magnitude and phase angle, and under these conditions there will still be no coupling between the series and the shunt branch arms. It is apparent that the behavior of the structure is similar to a bridge circuit, with the signal fed in one branch arm and the null detector on the other.

With matched loads on the two colinear arms, the impedance measured looking into either of the branch guides does not match the guide impedance. If, by the addition of some sort of matching elements, these impedances are made to match the wave-guide impedance, so that there are no standing waves in the input branch arms, the structure will possess additional qualities of balance.

As before, if power is fed in either branch arm and if the two colinear arms are terminated in matched impedances, the input power will divide equally between

these impedances, and there will be no coupling between the two branch arms. Because of the added matching structures, a matched impedance will be seen at the input. In addition, if a signal is fed in one of the colinear arms, it will divide equally between the two branch arms, if they are matched, and there will be no coupling to the other colinear arm. Also, the input impedance to the colinear arms under these conditions will match the guide impedance.

The matched structure is equivalent to a balanced four-terminal pair network, shown schematically in Fig. 36, in which each wave guide is represented by a terminal pair. In this equivalent network, if a signal is fed into any terminal pair and if the two adjacent terminal pairs are terminated in matched loads, the power will divide equally between these loads, and no voltage will appear across the opposite terminal pair.

The matching of the branch arms may be accomplished in a variety of ways. For maximum band width, it is desirable to have the matching structure as close as possible to the junction. One typical design is shown in Fig. 37, in which the H-plane tee is matched by the post and the E-plane tee by the asymmetrical inductive diaphragm.

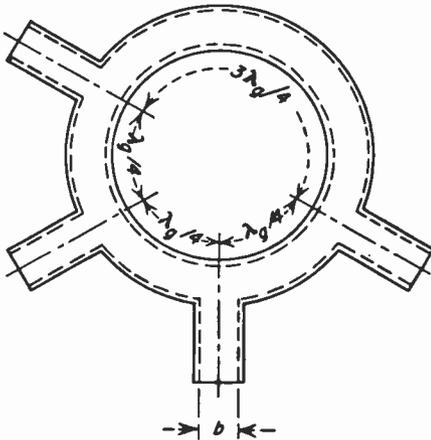


FIG. 39. Rat-race wave-guide bridge, electrically equivalent to magic tee of Fig. 35. The four branching guides connect with the ring in series tees. Distances between branch guides are measured along the mean circumference of the ring.

standing waves in a wave guide, the field strength in the guide must be measured as a function of position along the guide. The fields are completely contained within the outer conducting tube, so this tube must be slotted to allow entry of the probe or loop that samples the field. The slot should have a minimum effect upon the fields in the guide and should therefore be located where there is no current in the wave-guide walls transverse to the slot. For a rectangular wave guide operating in the $TE_{1,0}$ mode, the preferred location is in the center of the broad face of the guide, where the current has only an axial component (Fig. 40).

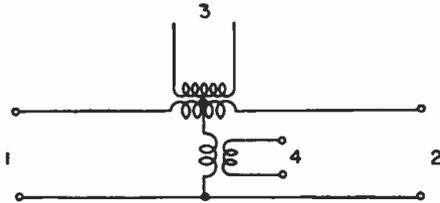


FIG. 38. Hybrid coil equivalent of magic tee of Fig. 35.

The l-f equivalent circuit of the magic tee is the hybrid coil, shown in Fig. 38, and the properties possessed by the two circuits are equivalent in many ways.

Several other structures possess properties electrically equivalent to the magic tee. One of these, known as a "ring" or "rat-race" bridge, is shown in Fig. 39. The distance between any two of the branch arms may be increased by an integral multiple number of half wavelengths without affecting the properties of this circuit. For reasons of impedance matching, it is desirable that the height of the wave guide forming the ring be somewhat less than the height of the four incoming guides. The band width of this structure is less than that of the magic tee, but the power-handling capacity is greater.

MISCELLANEOUS WAVE-GUIDE DEVICES

27. Standing-wave Detectors (Impedance Meters).

To measure the impedance meters. To measure the

A small fraction of the energy contained in the guide is coupled to a detector by a small probe or loop which extends into the guide through the slot. The coupling probe or loop is mounted on a carriage which moves along the guide, and the signal coupled to the detector depends upon the field strength at the location of the coupling element.

A coupling probe is usually preferred to a coupling loop because of the probe's mechanical simplicity, and because its position may be precisely determined. The probe should not extend into the guide far enough to disturb the field pattern inside the guide, and a penetration of 15 per cent or less of the guide height is considered good operating practice.¹ For this reason, accurate mechanical construction of standing-wave detectors is essential to operating accuracy, for slight variations in probe penetration will cause corresponding variations in the indicated field strength.

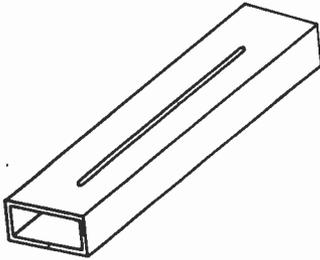


FIG. 40. Slot for measuring standing waves.

28. Matched Terminations for Wave Guide. Terminations for wave guides which present an impedance that matches the wave-guide impedance may be constructed in a variety of ways. The most successful designs, which remain matched over a considerable band of frequencies, utilize a section of wave guide to which some lossy or resistive element has been added to make it highly attenuating, with a tapered transition between the attenuating section and the incoming wave guide.

A relatively lossy dielectric filling the wave guide makes it highly attenuating. The phenolic resin-bonded plastics are satisfactory for this purpose, as are asbestos boards such as Transite which will handle higher powers. The tapered section may be constructed by tapering the dielectric; a taper of the type shown in Fig. 16 has proved satisfactory. If this taper is longer than 2λ , the design is very uncritical to either taper length or dielectric properties, provided a sufficient length of the high-loss dielectric is used in the attenuating section.

For lower power work, resistive films on a dielectric backing are very satisfactory. A platinized film on glass has been used, and carbon films on plastic board are very easy to work with. For maximum attenuation, the resistive film should lie in the plane of the electric field, in a region of strong field. For a rectangular guide in the $TE_{1,0}$ mode, the attenuation is high if the resistive film is placed parallel to the narrow faces of the guide, and centered with respect to the broad faces. The attenuation of this lossy section is matched to the input guide by tapering the film to a point as shown in Fig. 41.

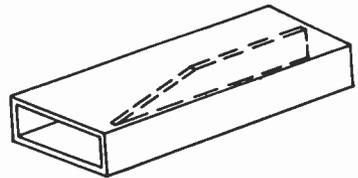


FIG. 41. Tapered strip of resistance film mounted inside rectangular guide to provide matched termination.

29. Fixed Attenuators.² Fixed attenuators for wave guides can be constructed in a manner similar to matched terminations. The amount of attenuation may be controlled by varying the length of the lossy section, as well as the attenuation per unit length. For fixed attenuators, it is usually necessary to match both ends of the lossy section to the wave guide, so there are tapers at each end of the lossy section. The tapers may be replaced by steps, similar to $\lambda/4$ matching sections. These are physically shorter but are matched over a reduced band width. The attenuation resulting from the use of lossy dielectrics is usually a function of the temperature of the dielectric, and also varies with its exposure to humidity. The resistive films are more

¹ Altar, W., F. B. Marshall, L. P. Hunter, Probe Error in Standing Wave Detectors, *Proc. IRE*, **34** (1), 33P, January, 1946.

² Gaffney, F. J., Microwave Measurements and Test Equipments, *Proc. IRE*, **34** (10), 775, October, 1946.

satisfactory in this respect. A typical fixed attenuator using a resistive film is shown in Fig. 42.

Directional couplers, discussed in a subsequent paragraph, are among the most satisfactory fixed attenuators.

30. Variable Attenuators. Variable attenuators used with wave guides fall into two classes: those utilizing the resistive attenuation of lossy dielectrics or resistance films, and those utilizing the reactive attenuation of below-cutoff wave guides. A typical resistive attenuator uses a resistive film of the same type used in matched terminations or fixed attenuators. The attenuation is varied by changing the position of the film within a wave guide, or by varying the amount of resistive film inserted through a slot in its outer wall.

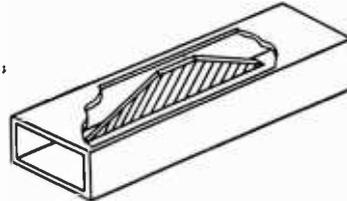


FIG. 42. Tapered resistance film to form fixed attenuator inside rectangular guide.

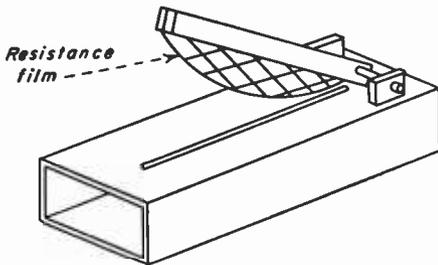


FIG. 43. Use of flap of resistance material as a variable attenuator by lowering it into a region of strong electric field.

An example of the latter type is shown in Fig. 43. These attenuators can be accurate and stable, if carefully designed, and have the distinct advantage that the minimum attenuation may be reduced essentially to zero by removing the film completely from the guide or by moving it to a region of negligible field strength. However, these attenuators require calibration against some standard, at the same or at a different frequency.

The reactive attenuation in a below-cutoff wave guide may be used in a variable attenuator,¹ the variation being accomplished by changing the length of

the below-cutoff wave guide. The attenuation in a below-cutoff wave guide may be calculated from the guide dimensions. For the $TE_{1,1}$ and $TM_{0,1}$ modes in round wave guide, the attenuation is given in Eqs. (16) and (17).

Typical attenuator designs, in which the below-cutoff wave guides are fed from coaxial lines, are shown in Fig. 44. The input impedance to a below-cutoff wave guide is a reactance, so if the input and output impedances of the attenuator are to match the line impedances, some lossy elements must be added. These may be simply added lengths of lossy lines, or they may be shunt resistance films, as shown schematically in Fig. 44.

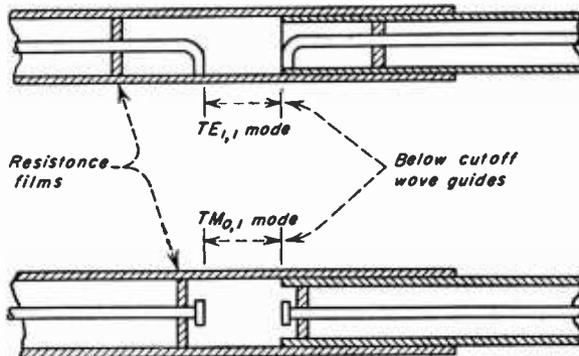


FIG. 44. Variation in below-cutoff sections as variable attenuators. Resistance films match the below-cutoff section to the input coaxial lines.

For the attenuation in the below-cutoff guides to equal the values given by the

¹ An example is given in H. R. Meahl, An Attenuator of S-band Energy, *Proc. IRE*, 35 (2), 211, February, 1947.

above formulas, the launching mechanism should excite only the desired mode in the below-cutoff wave guide. The attenuation in circular wave guide is lowest with the $TE_{1,1}$ mode, so if other modes are excited by the launching mechanism, they will attenuate more rapidly with distance, and the rate of attenuation will approach the theoretical value of the $TE_{1,1}$ mode at large attenuations. But attenuators using the $TM_{0,1}$ mode must be very carefully designed so that the undesired $TE_{1,1}$ mode will not be excited, otherwise the $TE_{1,1}$ mode will become increasingly strong relative to the desired $TM_{0,1}$ mode at large spacings and cause a departure from linearity.

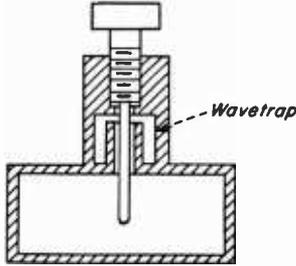


FIG. 45. Adjustable probe providing variable reactance shunting the guide. Wave trap gives electrical contact between probe and wave guide without metal-to-metal contact at inner surface of guide.

of matching any load impedance not a pure reactance to the characteristic impedance of a wave guide.

For tunable impedance-matching devices in wave guides, shorted stub lines of variable length may be used in various combinations, and the series or shunting reactance of the stub guides is varied by changing their lengths. Another type of variable reactance that is well suited to wave-guide structures is the adjustable probe, shown in Fig. 45. The shunt reactance of this probe is varied by changing its penetration into the guide. A typical impedance transformer may utilize two or three of these adjustable probes, fixed in position and spaced along the line. Another very useful design has the probe mounted on a sliding carriage and extending through a slot into the wave guide. With this arrangement, both the penetration and position of the probe may be continuously varied.

A special type of impedance transformer is the phase shifter, which is a section of a wave guide whose electrical length is continuously variable. A typical phase shifter, shown in Fig. 46, uses a tapered block of dielectric which partly fills the wave guide. The effect of the dielectric in a wave guide is to reduce the wave-guide wavelength, but the amount of reduction depends upon the electric field strength where the dielectric is located. So in the design of Fig. 46, as the dielectric block is moved toward the center of the wave guide to a region of greater electric field strength, the wavelength in that section of wave guide will be reduced, and the electrical length of the section of guide will correspondingly increase.

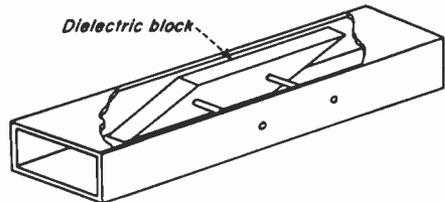


FIG. 46. Continuously variable phase shifter. The electrical length of the wave-guide section is increased by shifting the tapered dielectric block toward center of guide to region of greater electrical field.

Below-cutoff wave-guide attenuators are widely used where a precision variable attenuator is desired, as in signal generators, but their primary disadvantage is the large minimum attenuation. If the input and output coupling elements are brought too close together, the calibration will deviate greatly from theoretical values, and in addition the interaction between input and output will vary the impedances presented by the attenuator to the incoming line.

31. Impedance Transformers. Fixed impedance-transforming elements in wave guides frequently consist of the metallic diaphragms described on Sec. 18. These are equivalent to susceptances shunting the wave guide and are employed in the same fashion as stub transmission lines are used to tune conventional transmission-line systems; *i.e.*, if their size and position are correctly determined, these diaphragms are capable

32. Directional Couplers.¹ Directional couplers are measurement devices used in conjunction with h-f transmission lines, such as wave guides. They consist of two transmission lines, coupled together at an intersecting point or in an intersecting region. The coupling is in a manner such that an electromagnetic wave traveling in a single direction along one line will induce a wave traveling in a single direction along the second line. A wave traveling in the opposite direction on the second line will be induced by a wave traveling in the opposite direction along the first line.

The performance of a directional coupler may be described in terms of two quantities:

1. *Attenuation.* The attenuation of a coupler is the relative strength of the traveling wave on one line and the traveling wave which it induces on the second.
2. *Directivity.* Only in the ideal coupler does a single traveling wave on one line induce a wave traveling in a single direction on the second line. In a practical coupler, two waves traveling in opposite directions are induced. These are normally greatly unequal in strength, and their relative strength is the directivity of the coupler.

Directional couplers are widely used as fixed attenuators because the attenuation depends upon the coupling between lines. This coupling is generally through holes in the wall separating the two guides and is therefore quite unaffected by changes in temperature and humidity. Directional couplers also have application in the measurement of reflection coefficients and, in addition, permit a signal to be introduced in a single direction on a transmission line without actually breaking into the line.

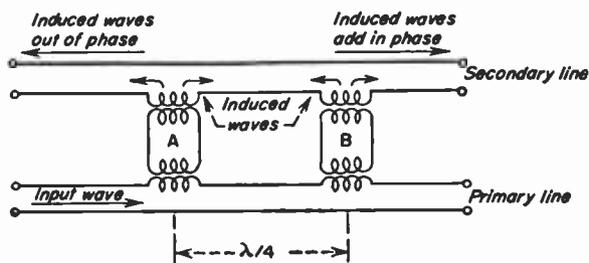


FIG. 47. Two-element directional coupler.

The schematic diagram of a typical directional coupler is shown in Fig. 47. The two transmission lines are weakly coupled together by two coupling links, spaced $\lambda/4$, as shown. A wave traveling from left to right on the primary line induces through each coupling link two waves traveling in opposite directions on the secondary line. Because of the spacing between coupling links in the primary line, link B will be excited 90 deg in phase behind link A. Because of the spacing between coupling links in the secondary line, the signal traveling to the right from link A will be retarded 90 deg with respect to the corresponding signal from link B, and the two waves will therefore reinforce each other. But the signal traveling to the left from link B will be retarded an additional 90 deg with respect to the corresponding signal from link A; the two signals will therefore be 180 deg out of phase and will cancel each other. As a result of these actions, the wave traveling from left to right on the primary line induces on the secondary line only a wave traveling from left to right. Correspondingly, a wave traveling from right to left on the secondary line is induced by a wave traveling in that direction on the primary line.

The coupling elements usually take the form of apertures through the walls of the adjacent wave guides, through which energy is coupled. A number of typical directional couplers are shown in Fig. 48. The two-hole coupler is the wave-guide analogue of the two-element coupler whose operation is described above. The three-hole coupler operates in a similar manner but has high directivity over a greater band of frequencies. The series-stub coupler is similar, but the coupling elements are sections of low-impedance wave guide. The single-hole coupler obtains its directive effect

¹ Mumford, W. W., Directional Couplers, *Proc. IRE*, **35** (2), 160, February, 1947. Early, H. C., A Wide-band Directional Coupler for Wave Guide, *Proc. IRE*, **34** (11), 883, November, 1946.

by having both electric and magnetic coupling through the hole, which reinforce in one direction and cancel in the other. The angle between the guides is chosen so that the two signals are equal in magnitude. The two-slot coupler has high directivity over a broad band because the 180-deg phase shift between the two coupled signals required for cancellation in one direction is accomplished by coupling to elements of

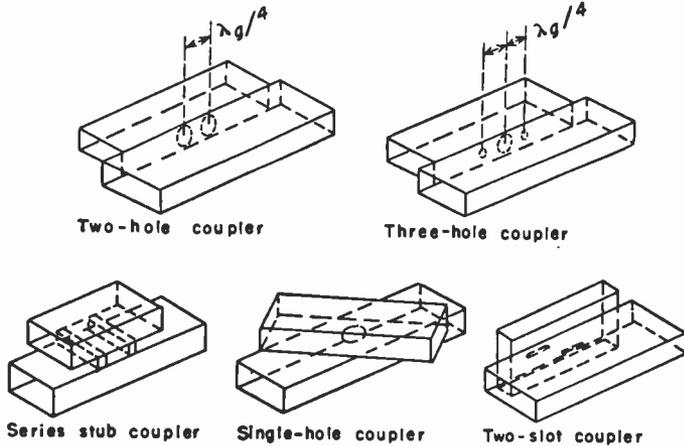


FIG. 48. Typical directional couplers for rectangular wave guide.

magnetic field in one guide that are 180 deg out of phase, this phase difference being independent of frequency.

CAVITY RESONATORS

Conventional resonant circuits, consisting of a coil and capacitor, are hardly suitable at microwave frequencies, as the physical dimensions become too small and the losses by radiation may be considerable. At higher radio frequencies, sections of transmission line are used as circuit elements, and coaxial lines offer the particular advantage of minimizing radiation losses because the electromagnetic fields are completely enclosed within the outer conductor.

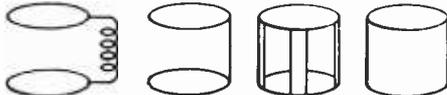


FIG. 49. Development of cavity resonator from a coil-capacitor circuit.

Coaxial line resonators are now generally included in the classification of cavity resonators, but the term did not come into general use until after it was realized that a hollow closed conducting box of arbitrary shape possessed electrical resonance properties similar to the conventional coil and capacitor circuit.¹ Rectangular, cylindrical, or spherical cavities may therefore be used at microwave frequencies to serve many of the same purposes that ordinary resonant circuits perform at lower frequencies, and in addition may possess an extraordinarily high efficiency. Q_s in the order of many thousand are not uncommon. Moreover, all cavity resonators possess not one but an infinite number of resonant frequencies, analogous to the infinite number of transmission modes of a wave guide.

An illustration of how a cylindrical cavity may be developed from a conventional coil and capacitor circuit is given in Fig. 49. The coil reduces to a single turn of wire at higher frequencies, this wire is in turn paralleled by others to reduce the inductance and increase the resonant frequency, until finally a solid wall connects the capacitor plates, and the cavity resonator is formed.

¹ Hansen, W. W., A Type of Electrical Resonator, *J. Applied Phys.*, 9 (10), 654, October, 1938.

With a cavity resonator, it is not possible in general to specify the exact inductance and capacitance of the equivalent resonant circuit. It is possible to specify, however, within limits of practical calculation the exact frequency of resonance for each of the resonance modes. Also the Q of a given cavity in a given mode may be uniquely defined. Finally, it is possible in many cases to specify the shunt impedance of a given cavity, once the path is determined along which the voltage shall be measured. In most instances, however, the choice of path is in some measure arbitrary.

33. Resonant Frequency. To calculate the resonant frequency of a cavity resonator, solutions to Maxwell's equations must be found which satisfy the boundary conditions imposed by the resonator. If the usual assumption is made that the cavity wall is a perfect conductor, the boundary conditions to be met are that no tangential electric field and no normal magnetic field exist at the surface of the cavity walls.

Under the above assumption, exact solutions are possible for resonators of simple geometrical shape. In practical cases, discrepancies between measured and calculated values for these simple shapes are extremely small, and usually attributable to mechanical imperfections in the cavity or to perturbations caused by the measuring device. A number of approximate methods of calculation have been developed which give more or less accurate answers for many more complex cavity shapes.

A principle of similitude may be applied to cavity resonators as follows: If all the linear dimensions of a cavity resonator are changed by a constant factor (and the resistivity of the walls changed by the same factor), the resonant wavelengths of all the normal modes in the cavity will be scaled by the same factor. The correction indicated in the parentheses is usually very small.

34. Resonator Q Values. The Q of a resonator in any given mode may be uniquely defined by the expression

$$Q = 2\pi \frac{\text{energy stored}}{\text{energy lost per cycle}} \tag{30}$$

If dielectric losses in the cavity resonator are neglected and if only losses resulting from conduction currents in the walls are considered, these losses are

$$\text{Energy lost per cycle} = \frac{\delta}{8} \int B^2 |d\sigma| \tag{31}$$

where δ = skin depth

B = magnetic field at wall of cavity

$d\sigma$ = element of area in cavity wall

The integral is carried out over the interior surface of the cavity. The energy stored in the cavity is

$$\text{Energy stored} = \frac{1}{8\pi} \int B^2 d\tau \tag{32}$$

with $d\tau$ an element of volume, and the integral carried out over the volume of the cavity. The Q of the cavity is then

$$Q = \frac{\lambda}{\delta} \frac{2 \int B^2 d\tau}{\lambda \int B^2 |d\sigma|} \tag{33}$$

It will be seen that Q is a dimensionless quantity. The above expression may be rewritten as

$$Q \frac{\delta}{\lambda} = \frac{2 \int B^2 d\tau}{\lambda \int B^2 |d\sigma|} \tag{34}$$

and the quantity $Q \frac{\delta}{\lambda}$ will be independent of frequency for a resonator of given shape and constant conductivity operating in a given mode. This quantity is sometimes known as the *form factor* of the cavity. The size of the resonator is inversely proportional to the frequency. As the skin depth varies as $\sqrt{\lambda}$, the Q of the cavity of given shape operating in a given mode will also vary as $\sqrt{\lambda}$, or inversely as \sqrt{f} .

For a nonreentrant resonator, the mean surface value of B^2 is approximately twice

the mean value throughout the volume, and it can be said approximately that

$$Q \frac{\delta}{\lambda} \cong \frac{1}{\lambda} \frac{V}{S} \tag{35}$$

where V is the volume and S the bounding surface of the resonator, given in the same linear units as λ . For highly reentrant resonators, it is a better approximation to multiply the Q as given by Eq. (35) by a factor approaching 2. The form factor increases with increasing volume-to-surface ratio, and therefore large cavities which operate in one of the higher modes of oscillation will generally have higher Q s than smaller cavities operating in simpler modes at the same wavelength. Cavities which are highly reentrant have a low volume-to-surface ratio and are likely to have lower Q s than simpler shapes will give.

35. Shunt Impedance of Cavity Resonators. The shunt resistance, or shunt impedance at resonance, of a cavity resonator may be defined by

$$R_{sh} = \frac{\text{voltage}^2}{2 \times \text{energy lost per sec}} \tag{36}$$

where the voltage is the line integral of electric field along some reasonable path, usually the path along which the maximum value of voltage is developed. If the resonator is being driven by a beam of electrons, the path that is chosen in a calculation of shunt impedance is the path followed by the electron beam. The shunt impedance of a typical, nonreentrant cavity at a frequency of 3,000 Mc will be in excess of 1 megohm. If the cavity is reentrant, the shunt impedance will be lower, but values in excess of 100,000 ohms are typical for highly reentrant cavities such as are used in velocity modulation tubes at these frequencies.

If the wall conductivity of a given cavity is increased, both Q and the shunt resistance of the cavity will increase proportionally to the square root of the wall conductivity. Also, if a cavity of given shape is scaled in size to resonate at a different frequency, both Q and the shunt resistance will be inversely proportional to \sqrt{f} .

To say that a cavity has a high Q does not necessarily imply, however, that it also has a high shunt resistance, nor does a high shunt resistance imply a correspondingly high Q .

It is possible to find cavity shapes in which a moderate Q is accompanied by an extremely high or an extremely low shunt resistance. A long cylindrical resonator operating in the $TM_{0,1,0}$ mode would have a high shunt resistance, and a reentrant coaxial resonator that is heavily capacitance loaded would have a low shunt resistance.

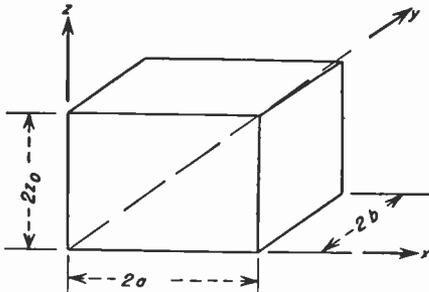


FIG. 50. Rectangular prism cavity resonator.

CHARACTERISTICS OF TYPICAL CAVITIES

36. Rectangular Resonators. The characteristics of a rectangular prism resonator, illustrated in Fig. 50, are readily calculated by analytical methods. A resonant wavelength will be found in such a resonator when

$$\lambda = \frac{4}{\sqrt{\left(\frac{l}{a}\right)^2 + \left(\frac{m}{b}\right)^2 + \left(\frac{n}{z_0}\right)^2}} \tag{37}$$

where l = number of half-wave variations of field along the x axis

m = number of half-wave variations of field along the y axis

n = number of half-wave variations of field along the z axis

$l, m, n = 0, 1, 2, 3, \dots$ but not more than one may equal zero for fields to exist

If the three dimensions of the resonator are equal and if the resonator is a perfect cube, there will be a twelvefold degeneracy; *i.e.*, 12 different modes will have identical resonant frequencies. If the three dimensions are unequal, this will be reduced to a twofold degeneracy, which in general will be destroyed by any irregularities in construction.

Counting each of the degenerate modes as a separate mode, the number of modes, dN , in a range $d\lambda$ is

$$dN = 8\pi \frac{V}{\lambda_1^4} d\lambda \tag{38}$$

where V is the resonator volume and λ_1 is the center of the wavelength band $d\lambda$. The number of resonant states, N , in a large rectangular prism resonator with resonant wavelengths greater than some minimum wavelength λ_2 is given approximately by

$$N = \frac{8\pi}{3} \frac{V}{\lambda_2^3} \tag{39}$$

This approximate formula is quite accurate even for low N .

For a resonator in which $a = b$ and where $l = m = 1$, and $n = 0$, the formulas for resonant wavelength, Q , and shunt impedance are given in Table 4. For large cubical resonators, operating in a high mode of oscillation, the Q is given approximately by

$$Q \frac{\delta}{\lambda} = \frac{a}{2\lambda_0} \tag{40}$$

37. Cylindrical Resonators.¹ For circular cylindrical resonators, as illustrated in Fig. 51, the infinite number of resonant modes that exist may be divided into two classes, corresponding to the two classes of transmission modes in a wave guide.

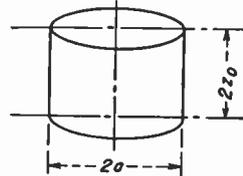


FIG. 51. Circular cylindrical cavity resonator.

1. *TE Modes.* The electric field in these modes is everywhere transverse to the axis, while the magnetic field has an axial component. The resonant wavelength of these modes is given by

$$\lambda = \frac{4}{\sqrt{\left(\frac{l}{z_0}\right)^2 + \left(\frac{2u_{m,n}'}{\pi a}\right)^2}} \tag{41}$$

Each of these modes occurs when the resonator is effectively a section of a circular wave guide that is an integral number of $\lambda/2$ long for some *TE* mode of transmission in the wave guide. The term l gives the number of $\lambda/2$ along the resonator axis and must therefore be an integral number. No modes exist in which $l = 0$. If the $TE_{m,n}$ mode in a wave guide is the mode being excited in the resonator, $u_{m,n}'$ is the n th root of the equation

$$J_m'(u') = 0 \tag{42}$$

Some of the lower roots of this equation are listed in Sec. 4.

The Q of the resonator for some of the lower modes may be obtained from Fig. 52.

The circular electric modes in circular cylindrical cavities, identified by subscripts of the form $TE_{0,n,l}$ are noteworthy for their exceptionally high Q . For this reason, they are ideal for high-precision cavity wavemeters, and also for echo boxes, as discussed in subsequent paragraphs.

2. *TM Modes.* The second class of resonant modes in circular cylindrical cavities are characterized by a magnetic field that is everywhere transverse to the axis, while

¹ Wilson, I. G., C. W. Schramm, J. P. Kinser, High Q Resonant Cavities for Microwave Testing, *Bell System Tech. J.*, 25 (3), 408, July, 1946. See also Kinser, J. P., and I. G. Wilson, End Plate and Side Wall Currents in Circular Cylindrical Cavity Resonator, *Bell System Tech. J.*, 26 (1), January 1947.

the electric field has an axial component. The resonant wavelength of these modes is given by

$$\lambda = \frac{4}{\sqrt{\left(\frac{l}{z_0}\right)^2 + \left(\frac{2u_{m,n}}{\pi a}\right)^2}} \quad (43)$$

As before, each of these modes occurs when the resonator is effectively a section of circular wave guide that is an integral number of $\lambda/2$ long, this time for some *TM*

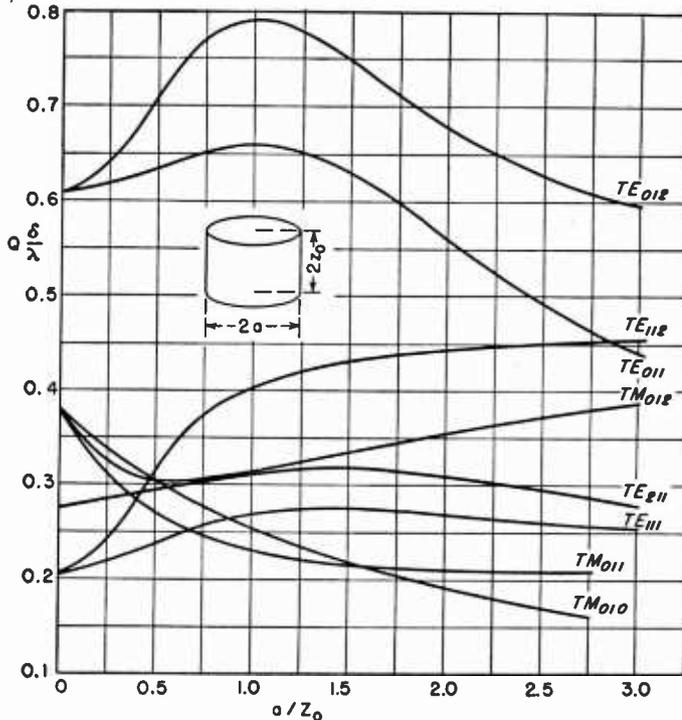


FIG. 52. Form factor of some of the lower modes in circular cylindrical resonators, from which Q may be determined.

mode of transmission in the wave guide. As before l gives the number of $\lambda/2$ along the axis and must therefore be an integral number, but in addition modes exist in which $l = 0$, corresponding to no transmission mode in a wave guide. These modes have an electric field that is everywhere parallel to the axis. If the $TM_{m,n}$ mode in a wave guide is the mode being excited in the resonator, $u_{m,n}$ is the n th root of the equation

$$J_m(u) = 0 \quad (44)$$

Some of the lower roots of this equation are listed in Sec. 4.

The simplest mode in a cylindrical cavity is the $TM_{0,1,0}$ mode. For this mode the characteristics are listed in Table 5.

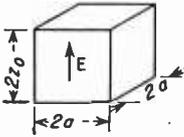
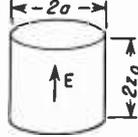
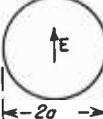
38. Spherical Resonators. The longest resonant wavelength will be found in a spherical cavity of radius a when

$$\lambda = 2.28a \quad (45)$$

and the next longest when

$$\lambda = 1.4a \quad (46)$$

Table 5. Characteristics of Simple Cavity Resonators Operating in the Lowest Mode of Oscillation

	Rectangular prism	Cylinder	Sphere
			
λ_0	2.828 a	2.61 a	2.28 a
$Q \frac{\delta}{\lambda}$	$0.353 \frac{1}{1 + \frac{a}{2z_0}}$	$0.383 \frac{1}{1 + \frac{a}{2z_0}}$	0.318
$R \frac{\delta}{\lambda}$	$120 \frac{z_0}{a} \frac{1}{1 + \frac{a}{2z_0}}$	$144 \frac{z_0}{a} \frac{1}{1 + \frac{a}{2z_0}}$	104.4

The field configuration for these two modes is shown in Fig. 53. Values for Q and shunt impedance of the first mode are given in Table 5.

39. Coaxial Line Resonators.¹ One type of coaxial line resonator is that shown in

Fig. 54. The lowest *TEM* resonance in this cavity will be found when the cavity is $\lambda/2$ long. This resonant wavelength is given by $\lambda = 4z_0$. The Q of the cavity in this mode is

$$Q \frac{\delta}{\lambda} = \frac{1}{4 + \frac{2z_0}{b} \log_e \frac{b}{a}} \tag{47}$$

and the highest Q is found when the diametric ratio is $b/a = 3.6$. The shunt impedance of this resonator is

$$R \frac{\delta}{\lambda} = \frac{60}{\pi} \frac{b}{z_0} \frac{\log_e^2 \frac{b}{a}}{1 + \frac{b}{a}} \frac{1}{1 + 2 \frac{b}{z_0} \frac{\log_e \frac{b}{a}}{1 + \frac{b}{a}}} \tag{48}$$

The maximum shunt impedance is found when $b/a = 9.2$. Highest Q and highest

¹ Barrow, W. L., and W. W. Mieher, Natural Oscillations of Electrical Cavity Resonators, *Proc. IRE*, 28 (4), 184, April, 1940. Some slightly misleading aspects of this paper are clarified by R. A. Kirkman and M. Kline, The Transverse Electric Modes in Coaxial Cavities, *Proc. IRE*, 34 (1), 33P, January, 1946.

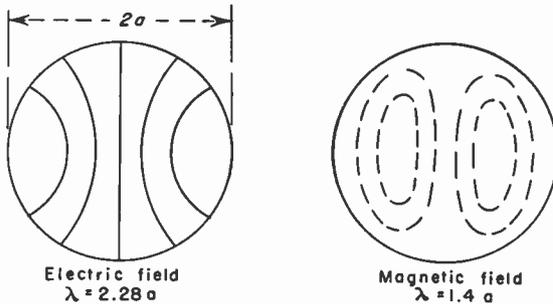


FIG. 53. Two lowest resonant modes in spherical cavity resonators.

shunt impedance require different values of diametric ratio, b/a . But neither shunt impedance nor Q is very critical to this ratio, as the Q will fall only to 78 per cent of its maximum value for $b/a = 9.2$, and the shunt impedance will fall only to 74 per cent of its maximum value for $b/a = 3.6$.

In addition to this lowest mode, resonant frequencies will be found whenever the cavity is an integral multiple number of $\lambda/2$ long, not only for the principal TEM mode, but also for any of the higher modes. None of these higher modes will give resonances unless the diameter of the line is sufficiently large. For the first of the higher modes to be above cutoff and be a possible source of resonances, the arithmetic mean circumference of the line should exceed a full wavelength.

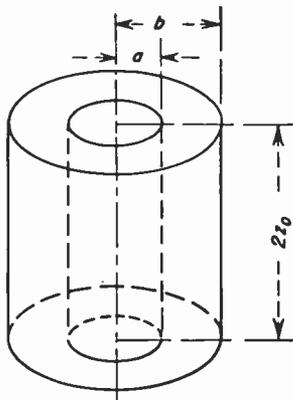


FIG. 54. Cavity resonator formed from a section of coaxial line shorted at both ends.

the cavity will resonate when the inner conductor is approximately $\lambda/4$ long. As the capacitance at the open end of the center conductor becomes appreciable, the length of the center conductor for resonance becomes less. An approximate calculation of resonant wavelength which neglects the fringing fields at the capacity gap gives for the resonant wavelength

$$\lambda = 2\pi \left(\frac{z_0 a^2}{2d} \log_e \frac{b}{a} \right)^{1/2} \quad (49)$$

The true resonant wavelength will always be greater than predicted by this formula, usually from 25 to 75 per cent greater. Accuracy will be best when the resonator is heavily capacitance-loaded.

For a reentrant resonator with considerable capacitance loading, the Q is given approximately by

$$Q \frac{\delta}{\lambda} = \frac{2z_0}{\lambda} \frac{\log_e b/a}{2 \log_e \frac{b}{a} + z_0 \left(\frac{1}{b} + \frac{1}{a} \right)} \quad (50)$$

where z_0 , b , and a are the cavity dimensions as given in Fig. 55. The shunt impedance of this cavity is approximately

¹ Examples of approximate techniques of calculation are given in the following papers: Hansen, W. W., On the Resonant Frequency of Closed Concentric Lines, *J. Applied Phys.*, 10 (1), 38, January, 1939. Mayer, E., Resonant Frequencies of the Nosed-in Cavity, *J. Applied Phys.*, 17 (12), 1046, December, 1946. Hahn, W. C., New Method for the Calculation of Cavity Resonators, *J. Applied Phys.*, 12 (1), 62, January, 1941. Mott, H., Calculation of the Electromagnetic Field, Frequency, and Circuit Parameters of High-frequency Resonator Cavities, *J. IEE*, 93, Pt. III (25), 335, September, 1946.

40. Capacitance-loaded Coaxial Line Resonators.¹ Resonators similar to the capacitance-loaded, coaxial line resonator shown in Fig. 55 are widely used in vacuum-tube u-h-f circuits. Klystron resonators are of this type in many instances, and disk-seal triodes are frequently employed with resonant circuits of this type.

Because of the relatively complex resonator geometry, accurate calculations of the resonant frequency and other characteristics are difficult to make, but approximate formulas are more readily developed.

Where the ratio of length to diameter of the cavity is large and the capacitance between the open end of the center conductor and the end of the cavity is small,

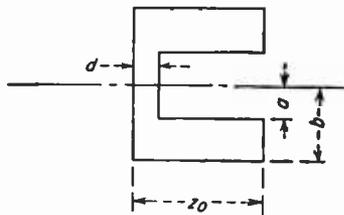


FIG. 55. Capacitance-loaded coaxial line resonator.

$$R \frac{\delta}{\lambda} = 60\pi \left(\frac{2z_0}{\lambda}\right)^2 \frac{\log_e^2 b/a}{2 \log_e \frac{b}{a} + z_0 \left(\frac{1}{b} + \frac{1}{a}\right)} \tag{51}$$

More exact calculations show that a resonator with fixed gap spacing d and inner radius a will have its maximum shunt impedance when the toroidal cross section is approximately square; *i.e.*, when $z_0 - d = b - a$.

If the cavity is relatively long and of small diameter, so that losses in the end plates may be neglected, but has appreciable loading capacitance, the Q is given approximately by

$$Q \frac{\delta}{\lambda} = \frac{\log_e \frac{b}{a}}{\pi\lambda \left(\frac{1}{b} + \frac{1}{a}\right)} \left(\frac{\sin \frac{4\pi z_0}{\lambda}}{\frac{z_0}{\lambda} + \frac{\sin \frac{4\pi z_0}{\lambda}}{4\pi}} \right) \tag{52}$$

and the shunt impedance approximately by

$$R \frac{\delta}{\lambda} = \frac{\log_e^2 b/a}{\lambda \left(\frac{1}{b} + \frac{1}{a}\right)} f(\lambda) \quad \text{ohms} \tag{53}$$

where $f(\lambda)$ is given in Fig. 56.

41. Coupling to Cavities. To utilize a cavity resonator, it is usually necessary to couple it to a transmission line. This may be done in a variety of ways. The coupling may be magnetic by means of a small coupling loop which links some of the magnetic flux lines of the desired mode in the resonator. The amount of coupling will be proportional to the magnetic flux enclosed by the loop and therefore to a first approximation proportional to the area of the loop.

To provide electric coupling to a cavity, the center conductor of a coaxial line may be extended as a probe into the cavity and will couple to the cavity if there is a component of electric field for the desired mode tangent to the probe. The coefficient of coupling will increase with increasing probe length and will also be greater if the probe is located at a point of high electric field strength.

Cavities may be coupled to wave guides by holes through the metal wall that separates the guide from the cavity. There will be magnetic coupling through a round hole if the magnetic field at the surface of the guide has a component parallel to the magnetic field at the adjoining surface of the cavity. There will be electric coupling if both guide and cavity have components of electric field normal to the common surface between them. The coefficient of coupling in both instances will be proportional to the third power of the hole diameter, if the wall is thin and the hole is small.

If a narrow slot is used instead of a round hole for coupling, the electric coupling will be small, but there may be appreciable magnetic coupling if there is a component of magnetic field in both guide and cavity parallel to the slot. Coupling between cavities and transmission lines by these various means is illustrated in Fig. 57.

When a cavity is coupled to a transmission line, the resonant frequency and Q of the cavity are affected by the coupled load, and there is usually some question as to just where the lines of demarcation are between the cavity, line, and coupling circuit.

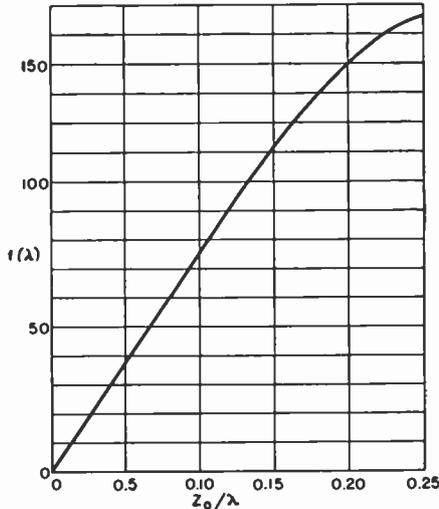


FIG. 56. The function $f(\lambda)$ used in Eq. (53) in the computation of shunt impedance of a capacitance-loaded coaxial line resonator.

For many purposes it is convenient to regard the resonant frequency of the cavity under these conditions as that frequency at which a minimum standing-wave ratio will be measured on the transmission line, when energy is fed into the cavity. This is the frequency at which free oscillations will occur in the circuit if the transmission line is terminated in its characteristic impedance.

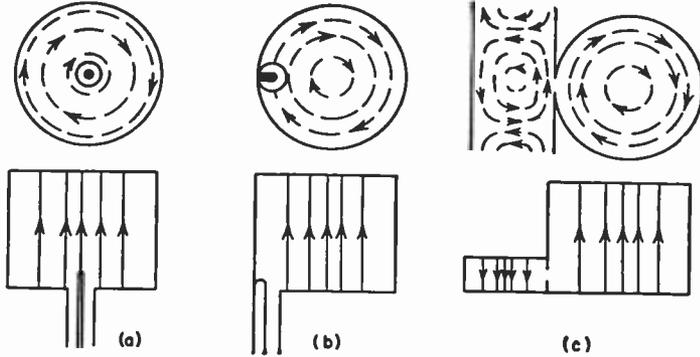


FIG. 57. Various methods of coupling between transmission lines and cavities. (a) Electric coupling between coaxial line and cavity by probe. (b) Magnetic coupling between coaxial line and cavity by loop. (c) Magnetic coupling between rectangular wave guide and cavity by a hole in the common wall.

The amount which the cavity is loaded by the coupled transmission line may be determined by the standing-wave ratio at resonance in the line feeding the cavity. The following Q s have been defined which are applicable to this situation:

$$\text{Unloaded } Q = Q_0$$

$$= \frac{2\pi f \times \text{energy stored in cavity}}{\text{power dissipated in cavity}}$$

$$\text{External } Q = Q_e$$

$$= \frac{2\pi f \times \text{energy stored in cavity}}{\text{power dissipated in load}}$$

$$\text{Loaded } Q = Q_L$$

$$= \frac{2\pi f \times \text{energy stored in cavity}}{\text{power dissipated in load and cavity}}$$

The various Q s are then related by the following equations:

$$\frac{Q_0}{Q_L} = 1 + \beta \tag{54}$$

$$\frac{Q_0}{Q_e} = \beta \tag{55}$$

$$\frac{Q_e}{Q_L} = 1 + \frac{1}{\beta} \tag{56}$$

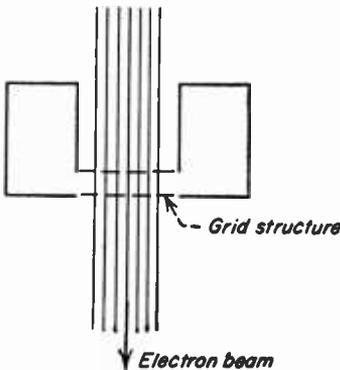


FIG. 58. An electron beam coupled to a cavity.

The term β is equal to the voltage standing-wave ratio at resonance in the input transmission line, if the positions of the minimums of the standing waves at resonance coincide with the positions of the maximums when the cavity is detuned. If the standing wave minimums at resonance coincide with the minimums when the cavity is detuned, β is the reciprocal of the standing-wave ratio measured at resonance on the input line.

Energy may be supplied to a cavity by an electron beam passing through the cavity. If the transit time through the resonator is short compared to a cycle, there will be power transferred from the beam to the cavity if the beam has an alternating component of current; *i.e.*, if the beam is in the form of bunches of electrons. These bunches

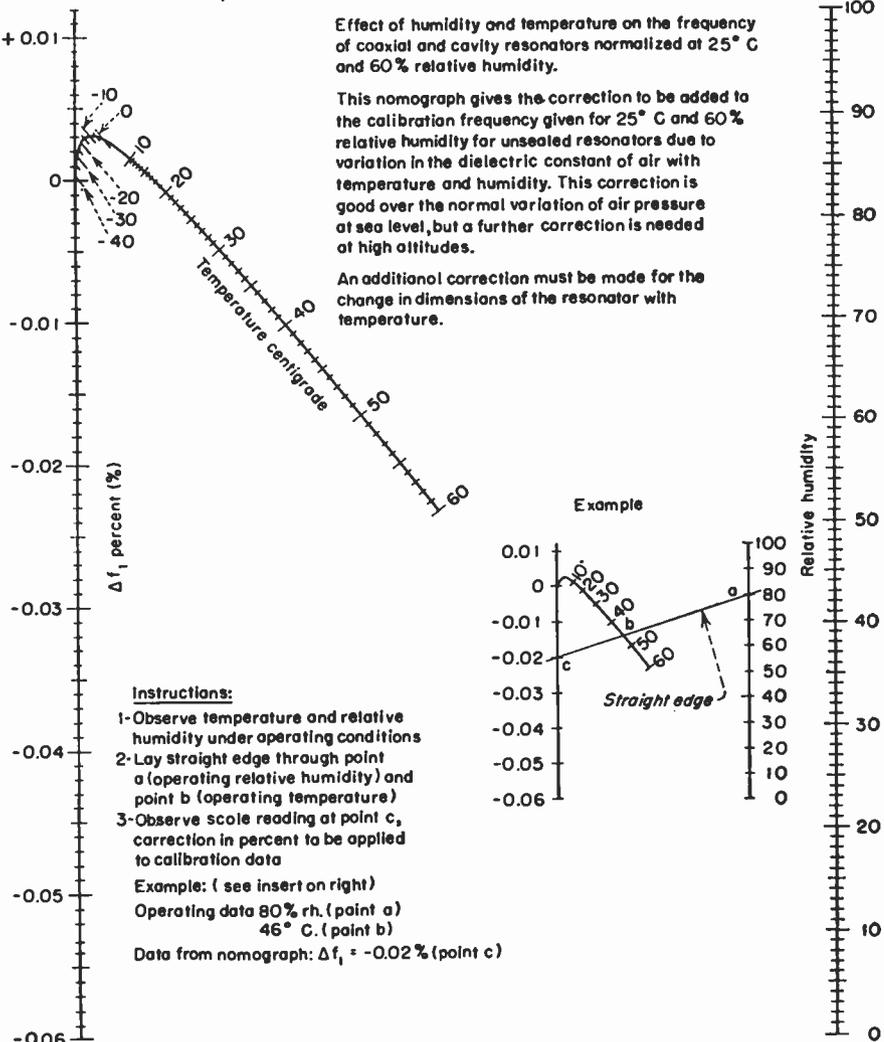


FIG. 59. Nomograph for calculating effect of humidity on resonant-line or cavity wave-meters.

may be formed either by velocity modulation action coupled with passage through a drift space, as in a klystron, or by the action of an alternating voltage applied to a control grid, as in a class C amplifier. The alternating voltage developed across the path of the beam will be equal to the product of resonator shunt impedance and the alternating component of beam current, provided the peak resonator voltage does not

exceed the beam voltage. A highly reentrant cavity, as shown in Fig. 58, is usually required to give an appreciable shunt impedance coupled with short transit time.

If the transit time through the cavity is appreciable compared to a cycle, calcula-

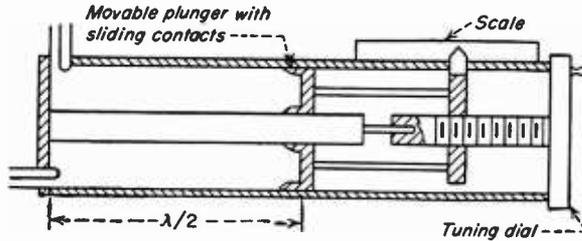


FIG. 60. Wavemeter cavity formed from section of coaxial line, shorted at both ends, of variable length.

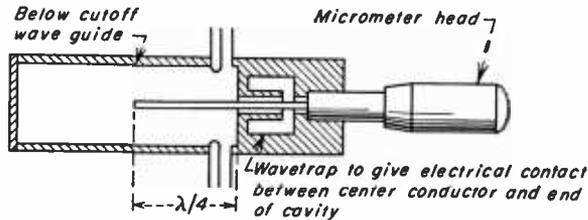


FIG. 61. Coaxial cavity wavemeter resonant when center conductor is approximately $\lambda/4$ long.

tion of the power transfer is more difficult. It can be shown that a beam which is unmodulated when entering a cavity is capable of exciting oscillations in the cavity if the transit time through the cavity is of the proper value.¹

42. Effect of Temperature and Humidity upon Resonant Frequency.

The resonant frequencies of a cavity depend upon the physical dimensions of the cavity, and if these dimensions vary with changing temperature, the resonant frequencies will also change. If the cavity is constructed of a single metal, the percentage change in resonant wavelength for any mode will be equal to the percentage change in linear dimensions, which in turn will depend upon the temperature change and the expansion coefficient of the metal.

The resonant frequencies of the cavity will also vary if the dielectric constant of the medium inside the cavity changes. If the cavity is not sealed, the changing dielectric constant of the air resulting from variations in temperature and humidity will cause changes in its resonant frequencies. These changes may be determined from the nomograph of Fig. 59.²

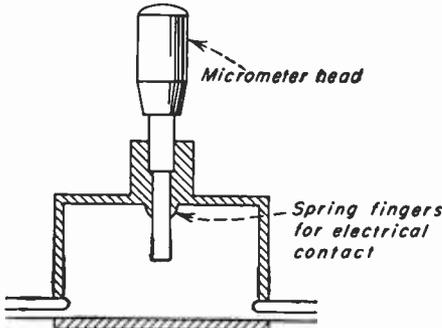


FIG. 62. Cavity wavemeter in which cavity is cross between $\lambda/4$ coaxial line resonator and a circular cylindrical resonator operating in the $TM_{0,0}$ mode.

¹ Marcum, J., Exchange of Energy between an Electron Beam and an Oscillating Electric Field. *J. Applied Phys.*, 17 (1), 4, January, 1946.

² Gaffney, F., Microwave Measurements and Test Equipments, *Proc. IRE*, 34 (10), October, 1946.

43. Cavity Wavemeters. The extremely high Q s associated with cavity resonators make them ideal for use as wavemeters at microwave frequencies. The simplest type of cavity wavemeter, which is really a coaxial line Lecher wire system, is shown in Fig. 60. The cavity is a section of coaxial line, shorted at both ends, which is resonant when $\lambda/2$ long. The resonant frequency is varied by moving the shorting plunger at one end of the cavity to change the length of the cavity.

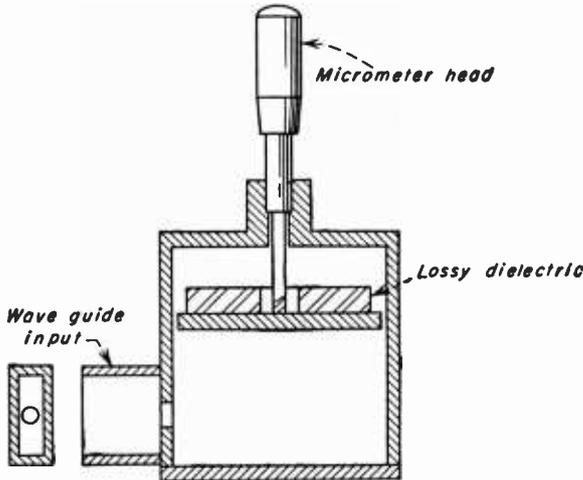


FIG. 63. High- Q cylindrical cavity wavemeter operating in $TE_{0,1,1}$ mode. The lossy dielectric behind the tunable end plate eliminates spurious resonances.

A second type of coaxial cavity that is widely used is shown in Fig. 61. This cavity is a $\lambda/4$ coaxial line, open at one end and shorted at the other. The resonant frequency is tuned by changing the length of the plunger forming the center conductor. A variation of this design is shown in Fig. 62.¹ This cavity is really a hybrid between a $\lambda/4$ coaxial line resonator and a cylindrical resonator operating in the $TM_{0,1,0}$ mode, resembling the former when the plunger is extended, and the latter when the plunger

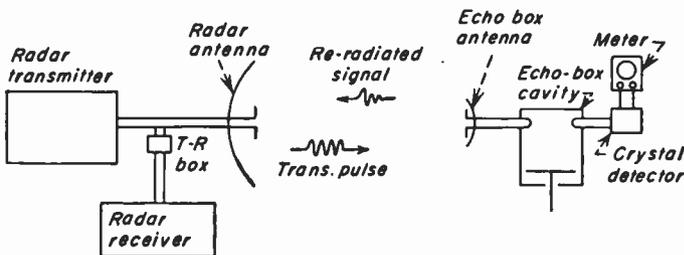


FIG. 64. Echo box being used for a check of the over-all performance of a radar system.

is nearly withdrawn. This design offers advantages of higher Q coupled with a tuning characteristic that is nearly linear in frequency.

The $TE_{1,0}$ mode in circular wave guide has an exceptionally low attenuation, and wavemeters employing this mode have a correspondingly high Q . A typical wavemeter design employing this mode is shown in Fig. 63. The resonant frequency is changed by moving the shorting plate at one end of the cavity. This plate is not

¹ Essen, L., Cavity-resonator Wavemeters, *Wireless Engr.*, 23 (272), 126, May, 1946.

required to make contact with the cavity walls, as with this mode of operation there is no current flowing across this gap from the side walls to the end plates. A cavity operating in this mode is capable of giving spurious indications by resonating in modes other than the desired $TE_{0,1,1}$ mode, and these spurious indications must be eliminated by damping out the undesired modes. This may be done with the slug of lossy dielectric shown mounted behind the tunable end plate in Fig. 63. For the undesired modes, currents will flow from the side walls to the end plate, and fields will penetrate into the space behind the tunable end plate. Energy is absorbed from these fields by the lossy dielectric, and the Q of the undesired modes is correspondingly lowered. Fields from the desired mode do not penetrate into this space, and the Q is therefore unaffected by the lossy dielectric.

44. Echo Boxes.¹ An echo box is a piece of test equipment used for over-all checks of radar systems. It consists of a high- Q cavity to which is loosely coupled a pickup antenna and also a crystal detector (Fig. 64). The echo box is placed in front of the antenna of the radar system, so that part of the energy radiated from the radar antenna during the transmitter pulse is picked up by the echo-box antenna and stored in the cavity in the form of oscillating electromagnetic fields. When the transmitter pulse is completed, this energy stored in the cavity will be reradiated through the echo-box antenna and detected by the radar receiver. The strength of this reradiated signal will decrease exponentially with time, and the length of time during which it can be detected by the radar receiver is a measure of the over-all performance of the radar system. This time, called the *ring time* of the echo box, depends upon the system tested.

For the rate of signal decay to be minimized, the Q of the echo-box cavity should be as high as possible. Two general types of cavities, tuned and untuned, are in use. The untuned cavity has dimensions large compared to a wavelength, and the various resonant modes are therefore spaced so closely in frequency that one or more will be excited by the frequency spectrum of the radar transmitter. For this reason it is unnecessary to tune the cavity to the frequency of the transmitter. A typical untuned echo box for a frequency of 3,000 Mc will be a nearly cubical box, approximately a meter on each side. A perfect cube is avoided to prevent mode degeneracy, as discussed in Sec. 36.

A tuned echo box is smaller and for that reason more convenient to use, but it must be tuned to the frequency of the radar system. To obtain the requisite high Q in a small cavity, the operating mode must be carefully chosen, and circular electric modes in circular wave guides are generally preferred. At 3,000 Mc, a cavity operating in the $TE_{0,1,1}$ mode has been found satisfactory with an operating Q over 40,000. At higher frequencies the ring time associated with a cavity of this design decreases rapidly, and it is therefore necessary to go to higher modes to get satisfactory performance. Much work has been done with cavities operating in the $TE_{0,1}$ mode in circular wave guide, but between 10 and 20 $\lambda/2$ long. These cavities give Q s well in excess of 100,000, but many troubles of interaction between the desired and undesired modes are encountered, which require very careful design to overcome.

References

- Brainerd, J. G., *et al.*: "Ultra-high-frequency Techniques," Van Nostrand, 1942.
 King, R. W. P., H. R. Mimno, and A. H. Wing: "Transmission Lines, Antennas, and Wave Guides," McGraw-Hill, 1945.
 Moreno, T.: "Microwave Transmission Design Data," McGraw-Hill, 1948.
 Ramo, S., and J. R. Whinnery: "Fields and Waves in Modern Radio," Wiley, 1944.
 Sarbacher, R. I., and W. A. Edson: "Hyper and Ultra High Frequency Engineering," Wiley, 1943.
 Schelkunoff, S. A.: "Electromagnetic Waves," Van Nostrand, 1943.
 Skilling, H. H.: "Fundamentals of Electric Waves," Wiley, 1942.
 Slater, J. C.: "Microwave Transmission," McGraw-Hill, 1942.
 Staff of the Radar School at M. I. T.: "Principles of Radar," McGraw-Hill, 1946.
 Young, V. J.: "Understanding Microwaves," Rider, 1946.

¹ Gaffney, *loc. cit.* Also Green, E. E., H. J. Fisher, and J. G. Ferguson, Techniques and Facilities for Microwave Radar Testing, *Bell System Tech. J.*, 26 (3), 435.

CHAPTER 16

LOUD-SPEAKERS AND ROOM ACOUSTICS

By HUGH S. KNOWLES¹

In the design and operation of electroacoustic devices, consideration must be given both to the physical or "objective" properties of the sounds that are to be reproduced and to the psychophysiological or "subjective" processes involved in hearing.

1. Sound is an alteration in pressure, particle displacement, or particle velocity propagated in an elastic material or the superposition of such propagated alterations.

2. Sound is also the sensation produced through the ear by the alterations described above. In case of possible confusion the term "sound wave" may be used for concept 1, and the term "sound sensation" for concept 2.²

In the case of a sound wave in air the pressure is alternately above and below atmospheric.

The velocity of propagation, c , of a sound wave of small amplitude is

$$c = 33,060 + 61\theta \quad \text{cm/sec} \quad (1)$$

where θ is the temperature in degrees centigrade. The wavelength λ is given by the relation $\lambda = c/f$, where f is the frequency in cycles per second. The density ρ of dry air at 20°C and at a pressure of 760 mm is 0.001205 g per cc.

The intensity of a plane or spherical "free" sound wave (no reflection) in the direction of propagation is

$$I = \frac{p^2}{\rho c} = 2.42 \times 10^{-9} p^2 \quad \text{watt/sq cm} \quad (2)$$

where p is the effective sound pressure (dynes per square centimeter).

The standard reference intensity is 10^{-16} watt per sq cm. The intensity level in decibels of a plane or spherical free sound wave in the direction of propagation is

$$L_I = 10 \log_{10} 2.42 \times 10^7 p^2 \quad (3)$$

In a plane or spherical free wave the intensity is proportional to the square of the pressure. In this case the pressure level in decibels of a sound wave is defined as

$$L_p = 20 \log_{10} p/p_0 = 74 + 20 \log_{10} p \quad (4)$$

where p_0 is the standard reference pressure of 0.0002 dyne per sq cm. Two pressures are said to differ by x db if 20 times the logarithm to the base 10 of their ratio is x . As in the analogous electrical case involving a field strength or voltage ratio, this is valid only if the impedances are identical so that the energy is proportional in both instances to the square of the respective pressures. For this reason Eq. (2), in general, does not hold in more complicated fields. On the other hand the use of Eq. (4) is justified in more complicated sound fields by the fact that the hearing sensation depends primarily on the sound pressure.

1. **Speech.** The variation in conversational speech power with frequency is shown in Fig. 1. The ratio of $\frac{1}{8}$ sec peak to averaged power in 15-sec intervals is roughly 20 db. In overloaded amplifiers such as are frequently used in public-address systems, the ratio may be 10 db or less. This ratio is important in temperature-limited loud-

¹ Jensen Mfg. Co.

² American Standard Z24.1, Acoustical Terminology.

speakers (see Tests). The distribution of energy with frequency is brought out differently in Fig. 2.

Articulation curves which give a measure of the "recognizability" of speech are shown in Fig. 3. The percentage of called sounds correctly recognized is the per cent articulation.

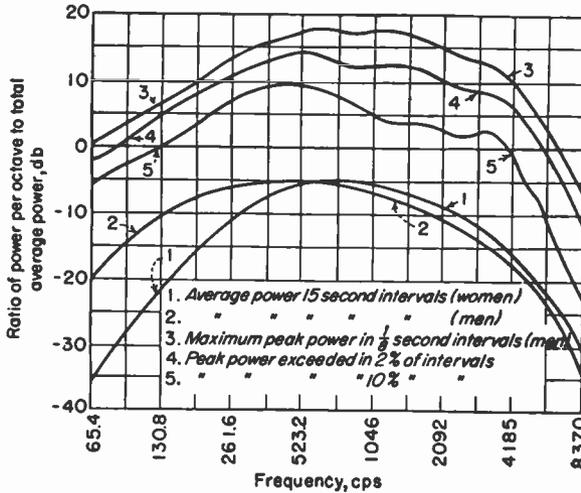


FIG. 1. Variation of conversational speech power with frequency. (After Sivian and Fletcher.)

Tests of syllable, sound, vowel, individual sound, and other types of articulation are now widely used in the laboratory and to an increasing extent in the field to determine the suitability of a system for the transmission of speech. "Intelligibility" tests, in which the content of a simple sentence is to be understood, are also used. On the average 30 per cent syllable articulation corresponds to nearly 90 per cent "discrete sentence" intelligibility, indicating the relative ease of understanding connected speech. From Figs. 2 and 3 we note that reproducing only the frequencies above 400 cycles halves the system power requirement and yet reduces the articulation by a negligible amount.

In a power-limited system in which speech articulation is important, the transmission band is sometimes limited to from 600 or 800 to 4,000 cycles, corresponding roughly to the 90 per cent articulation points at each end. This reduced band roughly quarters the power requirement.

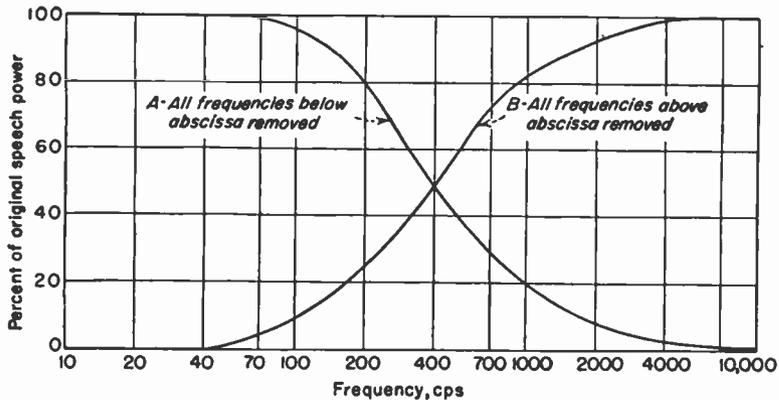


FIG. 2. Speech power variation with frequency. (After Fletcher.)

Articulation and naturalness are not to be confused. By successively raising the cutoff of high-pass filters and lowering the cutoff frequencies of low-pass filters, each by a barely perceptible amount, Schäfer has shown that the required transmission band for natural speech reproduction includes some 32 to 36 minimum perceptible

changes in band width.¹ The steps are roughly logarithmic. Some change in quality could be detected when frequencies above 8,000 were attenuated. The transmission of natural sounding speech and noises which accompany it therefore appears to require the transmission of all frequencies from 100 to about 10,000 cycles.

2. Music. The frequency distribution of the maximum and most probable peak powers for a 75-piece orchestra is shown in Fig. 4. The curves are based on average measurements of four selections which gave whole "spectrum" peak powers from 8 to 66 watts and average powers of 0.08 to 0.13 watt. Zero level corresponds to an average power of about 0.1 watt. As in the case of speech the average power over 15-sec intervals is about 1 per cent of the peak power in $\frac{1}{8}$ -sec intervals.

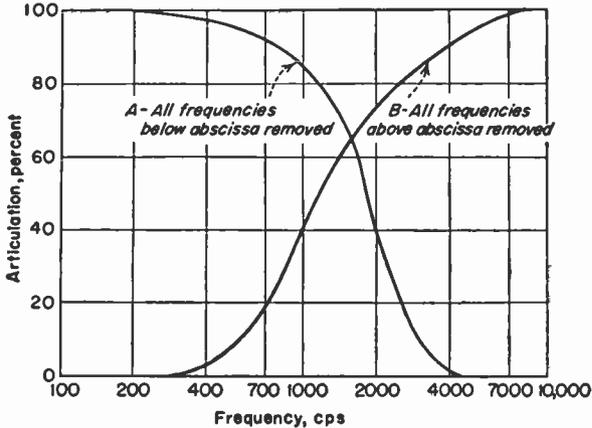


FIG. 3. Variation of articulation with transmitted frequency range. (After Fletcher.)

The power output of various musical instruments is shown in Table 1. The bass drum may radiate over a third of the peak power of a 75-piece orchestra. The large peaks in the 20- to 62.5-cycle range of the organ are well known to recorders and electronic organ people who find it desirable to use l-f stops which are "rich in harmonic development" and therefore sound much louder without badly overloading the record, amplifier, and speaker. The 15-in. cymbals follow the drums and organ closely in peak power output with 9.5 watts. Their maximum peaks occur in the 8,000- to 11,300-cycle range. Transmission systems having a "predistorted" frequency characteristic which includes a marked rise in h-f response in some part of the system

(such as f-m and television transmitters) are frequently overloaded by this instrument. The same problem occurs in recordings recorded with a similar characteristic. The high output of the trombone in the 2,000- to 2,800-cycle band near the frequency of maximum ear sensitivity gives the trombone (and other brass instruments) their piercing "bite." It has been found that the ear critically appraises the response

¹ Schäfer, E., The Audibility of Variations in Frequency Band in Speech Transmission, *Elek. Nachr.-tech.*, 15 (8), 237-240, August, 1938.

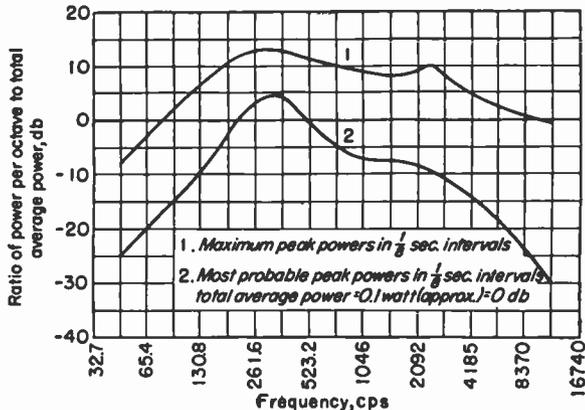


FIG. 4. Variation with frequency of the power output of a 75-piece orchestra. (After Fletcher.)

Table 1. Peak Powers in Music

Instrument	Microphone position and assumption in converting to total sound power	Field pressure, dynes per sq cm		Total peak power, watts	Percentage of intervals	Frequency band containing max peaks, cps
		Average in 15-sec interval	Peak in $\frac{1}{2}$ -sec interval			
Bass drum, 36 × 15 in..	3 ft in front, on axis. Radiation confined to a cylinder having drum diameter	99.0	1,260.0	24.6	6.0	250- 500
Bass drum, 30 × 12 in..	Same as above	35.0	980.0	13.4	1.0	125- 250
Snare drum.....	4 ft in front, 90 deg off axis. Peak pressure increased 8.5 db for 1-ft distance. Radiation confined to hemisphere	14.6	365.0	11.9	2.5	250- 500
15-in. cymbals.....	3-ft distance. Peak pressure increased 7.2 db for 1 ft. Radiation confined to hemisphere	18.0	360.0	9.5	7.5	8,000-11,300
Triangle.....	3-ft distance. Conversion as for cymbals	2.3	25.8	0.05	1.0	5,600- 8,000
Bass viol.....	3-ft distance. Radiation confined to hemisphere	4.2	37.8	0.156	2.0	62- 250
Bass saxophone.....	3-ft distance. Radiation confined to hemisphere	4.1	58.2	0.288	25.0	250- 500
BB $\frac{1}{2}$ tuba.....	3-ft distance. Conversion made from measurements with a complex sound source attached to a horn of similar size	5.4	43.2	0.206	17.0	250- 500
Trombone.....	3-ft distance. Conversion as for tuba	6.5	228.0	6.4	5.0	500- 700 2,000- 2,800
Trumpet.....	3-ft distance. Conversion as for tuba	8.6	54.2	0.314	18.0	250- 700
French horn.....	As for trumpet	3.8	27.0	0.053	6.0	250- 500
Clarinet.....	As for trumpet	3.3	26.4	0.050	5.5	250- 500
Flute.....	As for trumpet	1.6	25.6	0.055	1.0	700- 1,000 1,400- 2,000
Piccolo.....	As for trumpet	2.2	30.8	0.084	0.5	2,000- 2,800
Piano.....	10-ft distance. Room 29 × 29 × 13 ft. Reverberation time 1 sec 60-4,000 ~, average of 3 methods	2.6	23.4	0.267	16.0	250- 500
15-piece orchestra.....	6 ft from nearest instruments, in same room as piano. Average of 2 methods	7.9	126.0	9.0	1.5	250- 500 2,000- 2,800
75-piece orchestra.....	15 ft from nearest instrument in theater	4.6	129.0	66.5	1.0	250- 500 8,000-11,300
Pipe organ.....	Effective distance 15 ft. Radiation assumed uniform over $\frac{1}{4}$ sphere	20.0	90.0	12.6	36.0	20- 62.5

of a system in this range and that surprisingly small changes can be detected. This suggests that the balance of the brasses in a studio pickup merits special attention.

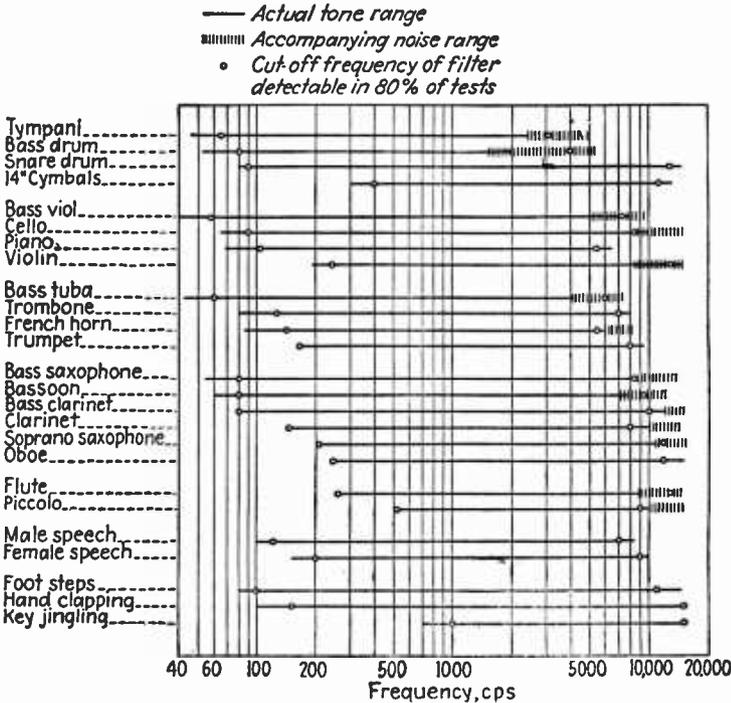


Fig. 5. Audible frequency ranges of some musical instruments and sounds. (After Snow.)

The audible frequency ranges of many musical instruments are shown in Fig. 5. The vertical ruled portions indicate the frequency range in which noises accompanying the playing of the instrument occur. While the elimination of these frequencies permits the fact that the frequency range is restricted to be detected, it does not mean that the quality is judged to be best with the unrestricted range. In many cases the quality of the reproduced music from instruments which radiate extraneous noises (reed, bowing, key, and others) is improved by eliminating the noise range.

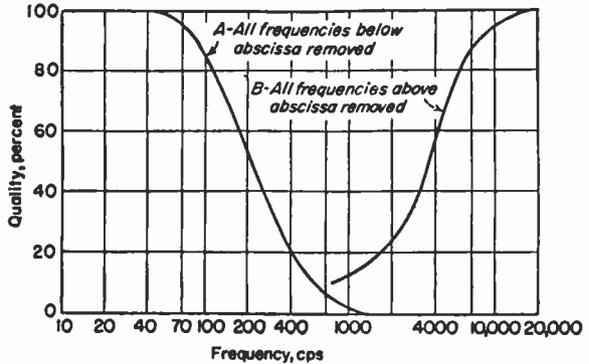


Fig. 6. Variation in quality of reproduced orchestral music with transmitted frequency range. (After Snow.)

In restricting the transmitted frequency range of reproduced music, we have to be

primarily concerned with the degradation in quality as judged by a good "sound jury" rather than with recognition of the selection played or the power distribution with frequency or "spectral composition" of the music. The average results of a test of this kind, using a jury of 10 and an 18-piece orchestra, are shown in Fig. 6. Considering the many variables involved, the maximum and mean deviations from the curve were surprisingly small. It was the judgment of the observers that the quality improved rapidly as the lower range was extended to 80 cycles and the upper to 8,000 cycles.

It has been found experimentally that, if the transmitted frequency range is to be restricted, good balance between low and high frequencies may be obtained by so choosing the range that approximately equal degradation in quality occurs because of loss of low and high frequencies. For reasonable degradation the product of these two frequencies is roughly 640,000. The square root of this product or the geometric mean of these frequencies is therefore roughly 800 cycles. A system transmitting more octaves below 800 cycles than above usually sounds "heavy," "thick," or "drummy." Likewise a system transmitting more octaves above 800 cycles than

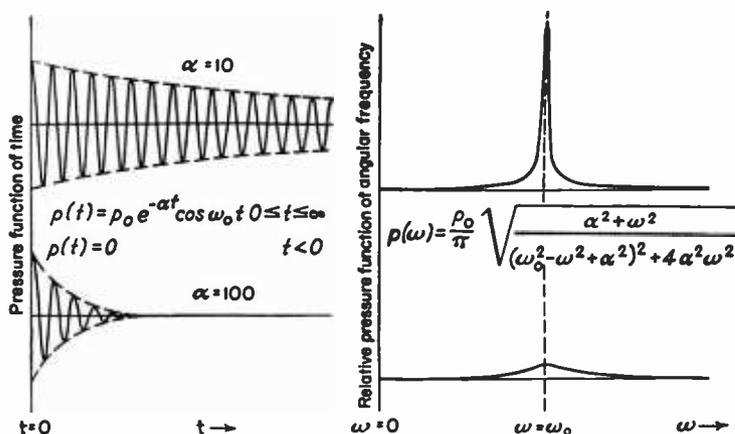


FIG. 7. Pressure variation with frequency for two isolated damped sinusoids with different rates of decay; initial amplitudes the same in the two sets of curves.

below will sound "thin" or "tinny." This assumes flat response in the range and similar cutoff characteristics. A sharp cutoff at one end will increase the apparent output at that end because of the transient response which accompanies such a cutoff. A peak in either range will increase the steady-state and transient response in that region. This can be only partly balanced by added response in the other range.

In considering the problem of reproducing sounds in a complete system including the effect of the room at the source of sound and at the source of the reproduced sound, it is desirable to know the energy distribution with frequency of a typical sound. The importance of this will be discussed under Room Acoustics. Since a common type of sound in music is a damped sinusoid, corresponding, for example, to the sound output of a plucked string instrument, the spectral analysis for two waves with different rates of decay is shown in Fig. 7. Any isolated wave train of this type contains energy which covers an infinite frequency interval. By analogy with the optical case the spectrum is said to be a continuous or band spectrum.

The highly damped wave contains appreciable energy at frequencies differing up to from 20 to 30 per cent from the frequency of a corresponding undamped wave. As the rate of decay is decreased, the wave train contains more energy, and an increasing amount of this is concentrated near the undamped frequency of the wave. In the

limit when the rate of decay is zero and the wave has existed for an infinite length of time, *i.e.*, when we have a steady state, the band spectrum degenerates into a line spectrum with all the energy concentrated at the undamped frequency of the wave.

The fact that music and speech are not of a steady-state character but vary from instant to instant (and therefore have a continuous distribution with frequency of their energy) substantially aids their satisfactory transmission in a room (see *Room Acoustics*).

3. Noise. Noise is an "unpitched" sound composed of a large number of discontinuous, nonperiodic sounds. Therefore the energy in noise is distributed in a continuous manner with frequency. A click, for example, closely approaches the hypothetical pulse which lasts for an infinitesimal length of time and the energy of which is continuously and uniformly distributed with frequency. A noise may have one or more broad peaks in its spectrum, but a sharp peak indicates a nearly periodic disturbance which will give the noise a definite pitch.

The properties of noises are of some importance because (1) the proper reproduction of intended noises may enhance the dramatic value of the reproduction, (2) the ambient noise levels in studios or halls and in rooms frequently limit the dynamic range at the "pickup" and "playback" points, and (3) they influence the response of the ear by producing masking or artificial deafening.

The reproduction of most noises requires the transmission of substantially the entire audible frequency range. For this reason noises are frequently used as test material in high-quality systems. The frequency ranges of footsteps, hand clapping, and key jingling are shown in Fig. 5. These indicate that it is particularly important that all the upper audible frequencies be transmitted.

The intensity level of various representative noises is listed in Table 2. In urban locations, particularly in large buildings, the ambient noise level in moderately quiet rooms is of the order of 45 to 60 db. This noise level is high enough so that even in specially treated broadcasting studios it frequently limits the dynamic range of the transmitter.

Even in relatively quiet residential sections the ambient noise level in a typical listening room is high enough so that it, too, places a lower limit on the intensity of the sound required to override the noise.

4. Hearing. All the data contained under Speech, Music, and Noise which were obtained by the use of a sound jury or listener involve the sensation produced in the listener by the designated source of sound. All tests of this type depend to some extent on the techniques employed and, of course, on the observers. All similar tests are of principal value when the jury is composed of a large selected sample with known hearing characteristics.

One of the most important properties of sound is its loudness. This has been found to vary with both the frequency and intensity of the sound. To a rough approximation it has been found that in the middle frequency range equal percentage increases in intensity produce equal increases in loudness. The *loudness* is the magnitude of the hearing sensation and is assumed proportional to the number of nerve impulses reaching the brain per second.

The *loudness level* contours for a sample of 200 ears are shown in Fig. 8. These curves were obtained by alternately listening to a sound of arbitrary frequency and intensity and comparing it with a 1,000-cycle tone the intensity of which was adjusted until the two were judged to be equally loud. At 1,000 cycles therefore the loudness level of the sound corresponds to the intensity level because this is the reference test frequency. The intensity is that which exists in an undisturbed sound field before the listener is immersed in it. The observer faces the source and listens to the sound binaurally. By plotting the differences in minimum audible field intensities for sound of normal and random incidence found by Sivian, we obtain the dotted curve in Fig. 8. This indicates that the other contours for sound of random incidence would also be more regular.

Recently reports have been made by Beasley on a sample of 16,000 ears. Some of

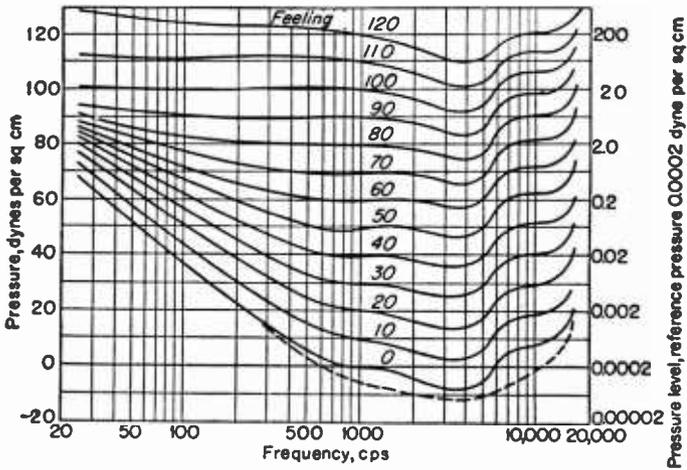


FIG. 8. Loudness level curves showing variation in sound intensity with frequency required to produce a sound judged to be as loud as the 1,000-cycle reference sound intensity given on the curves. (After Fletcher and Munson.) Solid curves obtained with listener facing sound source. Dashed curve indicates threshold (corresponding to solid curve 0) but for sound of random incidence. (After Sivian.)

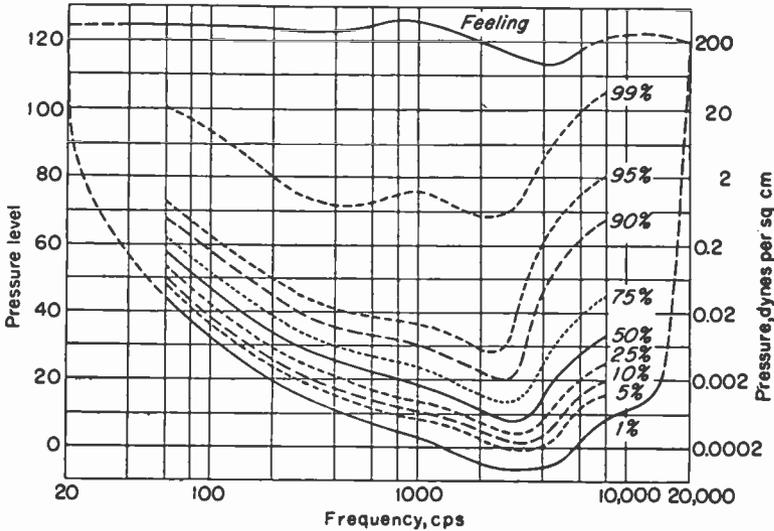


FIG. 9. Threshold of hearing curves for large population sample. Percentage figures indicate percentage of sample tested having a hearing threshold lower than the corresponding curve. (After Beasley.)

the results are shown in Fig. 9. The curves show the percentage of the sample tested which had lower thresholds of hearing than the indicated value. For example, the solid curve marked 50 per cent indicates that 50 per cent of the ears tested had thresholds of hearing lower than that indicated by this curve. From these data we see that the Fletcher and Munson threshold curves are for ears in the best 1 per cent

of the 16,000-car sample, and that hearing deficiencies are prevalent enough to justify their consideration in equipment design.

Table 2. Noise Levels

Pressure, dynes per sq cm	Noise level, db above reference threshold	Type of noise
630	130	Pain threshold
250	122	Airplane—1,600 rpm, 18 ft
45	109	Boiler factory
25	102	Subway train passing station
13	96	Elevated train—15 ft
4.0	86	Heavy traffic—15 ft
2.0	80	Average truck—15 ft
1.3	75	Average factory location
6.3×10^{-1}	70	Average automobile—15 ft
3.2×10^{-1}	64	Department store
1.1×10^{-1}	55	Average office
2.8×10^{-2}	43	Quiet office
6.3×10^{-3}	30	Very quiet residence
1.4×10^{-3}	17	Gentle whisper—5 ft
4.5×10^{-4}	7	Threshold (for street noise)

The loudness or apparent response or transmission characteristic of a system emitting a plane free sound wave of three constant intensity levels is shown in Fig. 10. A loud sound (constant 100-db intensity level) seems almost equally loud from 30 to 6,000 cycles. A sound of moderate intensity (constant 60-db intensity level) is inaudible below 60 cycles and increases in loudness rapidly up to 400 cycles. In the presence of noise, masking would substantially reduce the loudness at low intensities. The l-f characteristic varies much more rapidly with intensity level than the high and for this reason compensated volume controls are designed to have their maximum effect at low frequencies. compensated characteristic is a purely

It should be noted that the intensity-level com- arbitrary thing and, although it is considered

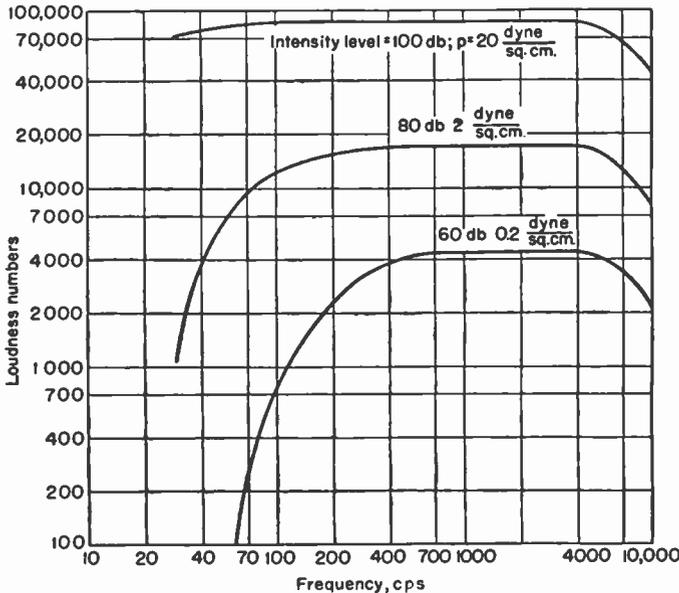


FIG. 10. Loudness variation with frequency for three pure tones of the indicated intensity showing reduced loudness of low-intensity l-f sounds.

superior by some, it does not correspond to our normal experience. In practice, when we go some distance from the source, the low frequencies drop out. When an orchestra plays at low intensity, we get the same effect.

The effect of noise on hearing is to produce artificial deafness or "masking." The nature of the apparent deafness which results depends on the spectral composition of the noise. Many noises produce fairly uniform deafening or masking. The effect of moderate noise levels is to decrease articulation. This may be largely compensated by raising the intensity level of the sound.

LOUD-SPEAKERS

A loud-speaker is a device which is actuated by electrical signal energy and radiates acoustical energy into a room or open air. The shorter term *speaker* is used when no confusion with a person addressing a microphone results.

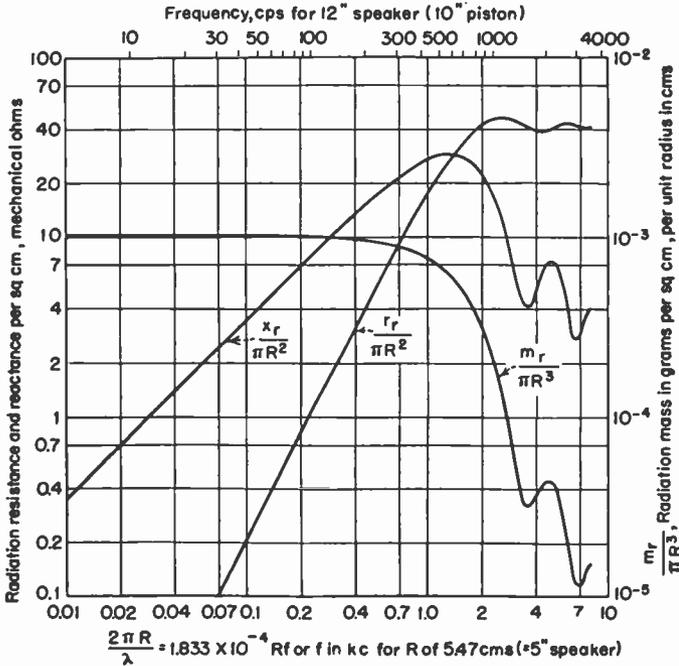


FIG. 11. Radiation resistance, reactance, and mass per square centimeter of a flat, rigid piston vibrating in an infinite, rigid nonabsorbing baffle. Piston radiates into a solid angle, $\Omega = 2\pi$ steradians (hemisphere).

The selection and installation of a speaker as well as its design should be guided by the problem of coupling an electrical signal source as efficiently as possible to an acoustical load. This involves the determination of the acoustical load or radiation impedance and selection of a diaphragm, motor, and means for coupling the loaded loud-speaker to an electrical signal source. The performance of the speaker is intimately dependent upon the nature of its acoustic load and should not be considered apart from it. The nature of the radiating system, and therefore the acoustic load impedance it sees, is primarily determined by space, acoustical environment, and cost factors.

5. Radiation Impedance. When a vibrating diaphragm is placed in contact with air, its impedance to motion is altered. The added impedance seen by the surfaces which emit useful sound energy may be called the *radiation impedance*. By analogy with antenna systems the resistive part is called the *radiation resistance*. The radia-

tion reactance or reactive part is usually positive, and the corresponding apparent mass may be called the *radiation mass*. The radiation impedance seen by a diaphragm depends on its size, shape, the frequency, the acoustical environment, and the medium into which it radiates.

6. Single Piston. The average radiation impedance per unit area seen by a flat circular piston vibrating in a thin, rigid, nonabsorbent, infinite plane or baffle in air is shown in Fig. 11. When the length of the radiated sound wave λ exceeds the circumference of the piston, $2\pi R$, the radiation resistance is nearly proportional to the square of the frequency. This relation between the wavelength and piston circumference corresponds to the frequency, being less than 4,310 divided by the diameter in inches. In this frequency range the piston *velocity* should vary inversely with frequency to radiate constant power since this is equal to the square of the rms velocity and the radiation resistance [see Eq. (5)]. This variation in velocity with frequency is usually obtained by placing the fundamental resonant frequency of the diaphragm and motor near the lowest frequency to be transmitted so the system has mass reactance or is "mass-controlled" in this frequency range. When the frequency exceeds twice this value, the resistance is very nearly 41.4 mechanical ohms per square centimeter and the diaphragm (real or virtual) is efficiently coupled to the air (see Diaphragms, Size).

When the length of the radiated wave exceeds the circumference of the piston, the air increases the apparent mass of each side of the diaphragm by approximately the mass of air contained in a cylinder whose base is the piston and whose height is 0.85 times the piston radius. At high frequencies the radiation mass ("accession to inertia") and the mass reactance decrease and approach zero for infinite frequency.

7. Mutual Radiation Impedance. When a sound wave radiated from one surface of a diaphragm has access to another surface of the same diaphragm or to a surface of another diaphragm, there is said to be coupling between the surfaces. Consideration of this mutual radiation impedance is simplified by fixing attention on what occurs

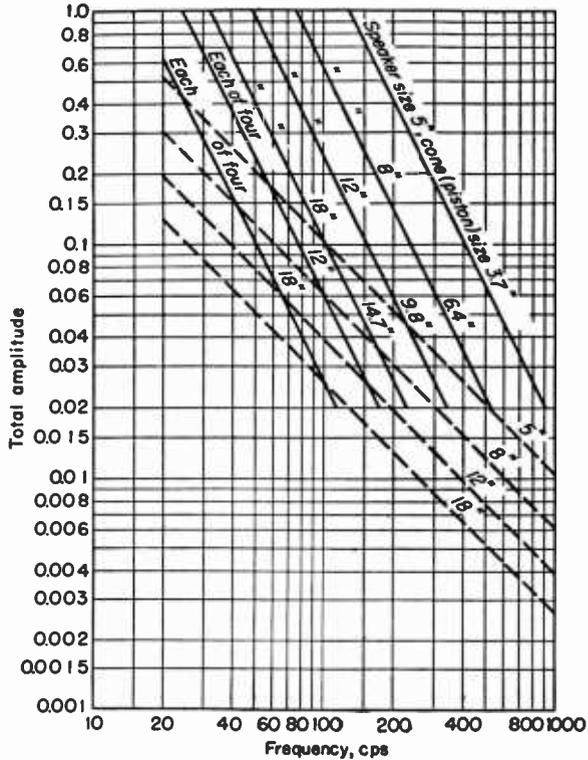


FIG. 12. Total displacement required of diaphragm of indicated size to radiate 1 watt. Solid curves for pistons radiating into a hemisphere at low frequencies where the radiation resistance is proportional to the square of the frequency ($2\pi R/\lambda$ less than about 1.4 in Fig. 11). Dashed curves for constant radiation resistance of 41.5 mechanical ohms per square centimeter (exponential horn value well above horn cutoff frequency).

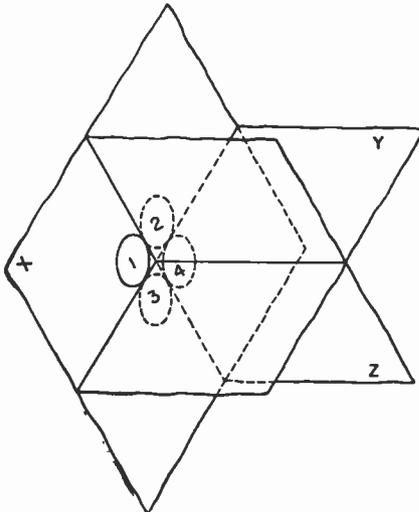


FIG. 13. Primary images 2, 3, and 4 of piston 1 introduced by planes Y and Z.

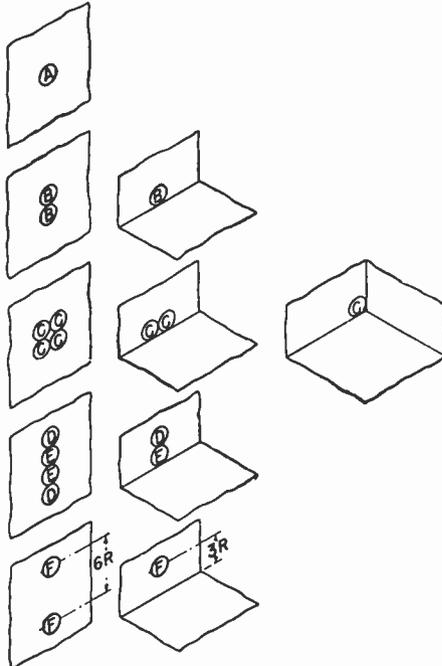


FIG. 14. Effect of adding pistons and reflecting planes on radiation impedance. All pistons marked with the same letter see the same radiation impedance.

at each diaphragm. The motion of the diaphragm is opposed by the ("self-") radiation impedance. It is also opposed or aided by the force exerted on it by the waves generated by any other diaphragms which are coupled to it. The (complex) ratio of the force due to all other diaphragms to the velocity of the diaphragm itself is the mechanical impedance seen by the diaphragm due to the other diaphragms. This we will call the total mutual radiation impedance.

The total radiation impedance seen by a diaphragm is the sum of the self- and the "mutual" radiation impedances. The acoustic power P_a radiated by a diaphragm is

$$P_a = (r_s + r_M)v^2 \times 10^{-7} \text{ watt} \quad (5)$$

where r_s = real part of self-radiation impedance (total)

r_M = real part of mutual-radiation impedance (total)

v = rms diaphragm velocity, cm per sec

Note that the velocity appears as current would in the corresponding electrical equation. The diaphragm displacement is $v/2\pi f$, where f is the frequency. The total displacement of various diaphragms required to radiate 1 watt is shown in Fig. 12. These curves clearly show the need for large diaphragms if appreciable 1-f power is to be radiated.

By knowing the self- and mutual radiation impedances of diaphragms mounted in a single infinite baffle, we can determine the impedance seen when other baffles are added. In Fig. 13 assume four pistons (1, 2, 3, 4) mounted in the X plane. Because of symmetry there is no net sound flux through the plane Y since for every positive vertical (z) component from pistons 3, 4 there is a negative component downward from pistons 1, 2. We may therefore introduce the rigid, thin, nonabsorbent plane or baffle Y without altering the impedance seen by any of the pistons. With Y in place we may remove pistons 3, 4, and pistons 1, 2 will continue to see the same impedance. The sound wave reflected by the plane corresponds

exactly to the wave which would come from diaphragms 3, 4 and therefore the plane is said to have created "primary images" (by analogy with the optical case) of diaphragms 1, 2 which 1, 2 cannot distinguish from the real diaphragms 3, 4. Similarly the plane *Z* may be introduced and pistons 2, 3, 4 removed, leaving 1 looking into its original impedance. In all cases pistons of equal size, vibrating in phase and with the same amplitude in infinite, rigid nonabsorbing baffles, are assumed. The relations hold approximately when the baffles are a wavelength or more long. Finite impedance of a baffle may be treated by assuming reduced amplitude of the image to account for absorption and a change in phase to account for the reactive part of the impedance. The principle is readily extended to multiple sources of arbitrary size, phase, and displacement such as occur in vented enclosures, labyrinths, and the like.

Several piston combinations are shown in Fig. 14. All pistons marked with the same letter see the same radiation impedance. The ratio of the radiation resistance

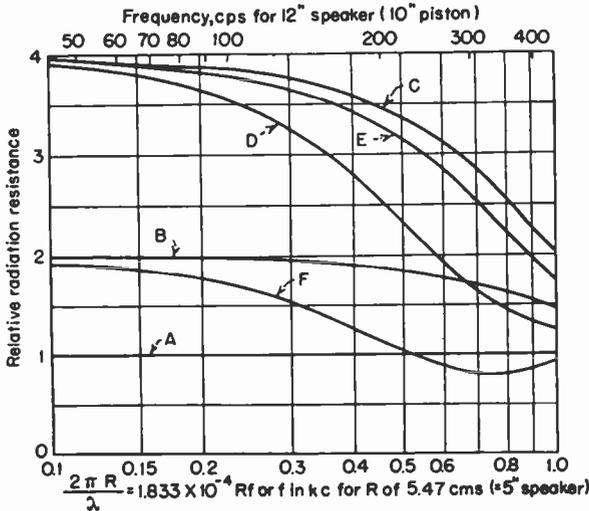


FIG. 15. Ratio of radiation resistance seen by a piston in the presence of one or more others (real or images) vibrating with equal amplitude and phase to radiation resistance it would see alone radiating into a hemisphere. Designating letter of curve is same as that of corresponding piston in Fig. 14. (After Klapman.)

and reactance seen in each case to that seen by a single piston *A* is shown in Figs. 15 and 16. The actual impedance is therefore obtained by multiplying the ordinates of Fig. 15 or 16 by the corresponding ordinate of Fig. 11. The letters on the curves correspond to those on the pistons in Fig. 14.

Values of the ordinate less than 1 indicate the piston sees less resistance or reactance than it would if alone in a single infinite plane. This occurs when the time delay in the wave from one diaphragm and the frequency are such as to give out-of-phase components at the other.

Figure 15 shows that the radiation resistance is increased by the largest factor and over the greatest frequency range when each diaphragm is as close to all others as is possible. At low frequencies the group then behaves as a single large piston. Figures 15 and 16 show that the radiation resistance increases much more rapidly than the reactance as speakers (or their images) are added. The exact improvement in efficiency which results depends on the impedance seen looking back into the speaker diaphragm but a typical speaker efficiency is increased by a factor of nearly 2 (3 db) for case *B* and 3.2 (about 5 db) for case *C*. This indicates qualitatively the improve-

ment gained by operating a speaker at the intersection of the floor and wall and in a corner, respectively.

For a given diaphragm amplitude one speaker in locations *B* and *C* will radiate two and four times as much l-f energy, respectively, as one in location *A*. The radiated power for constant amplitude is therefore proportional to the square of the number of actual diaphragms. The radiation resistance at high frequencies is not improved by the use of additional speakers. A group of speakers therefore has better low but no better high response than a single one, and they therefore sound as though they had relatively less high response. Except for cost reasons multiple speakers are usually preferred to a single speaker with the same size motor because (1) the small diaphragms are lighter per unit area than a large one of adequate rigidity, giving better efficiency

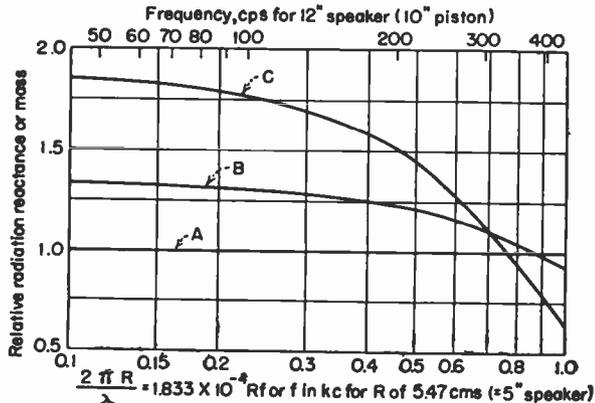


FIG. 16. Ratio of radiation reactance (or mass) seen by a piston in the presence of one or more others (real or images) vibrating with equal amplitude and phase to reactance (or mass) it would see alone radiating into a hemisphere. Designating letter of curve is same as that of corresponding piston in Fig. 14. (After Klapman.)

and high response; (2) the angle of individual speakers may be adjusted a moderate amount to give a good h-f directional pattern without injuring the l-f response; (3) improved reliability, since failure of a single unit usually does not seriously affect the performance of the group; and (4) the temperature rise of each voice coil is reduced.

MULTIPLE LOUD-SPEAKERS

Some of the numerous advantages of multiple direct-radiator speakers where these all cover the same frequency range are discussed in Sec. 7, Mutual Radiation Impedance, above. Multiple-speaker systems in which the speakers cover complementary frequency ranges also have certain advantages and are widely used. The more important advantages are (1) improved frequency response, since each type of unit covers a moderate range; (2) higher system efficiency, for the same reason; (3) improved directivity characteristic, since the diaphragm (or horn mouth) for the highest frequency range may be made relatively small (see Figs. 17, 18, and 19); (4) improved transient response, since many of the artifices used to obtain extended frequency ranges in single units make the transient response worse, particularly at high frequencies; (5) reduced intermodulation, since large amplitudes are confined to speaker reproducing low frequencies; and (6) reduced frequency modulation which occurs when a single diaphragm moves with large amplitude with respect to the listener, thereby altering the frequency (due to the Doppler effect).

8. Piston Directivity. With rising frequency the radiation from a rigid piston becomes increasingly concentrated on the axis, as shown in Figs. 17 and 18. Figure 17

shows that, if the flat axial pressure response curve so often sought after is obtained, the total radiation and therefore the efficiency are actually falling rapidly at high frequencies. Conversely, if a speaker is to have constant efficiency its axial pressure

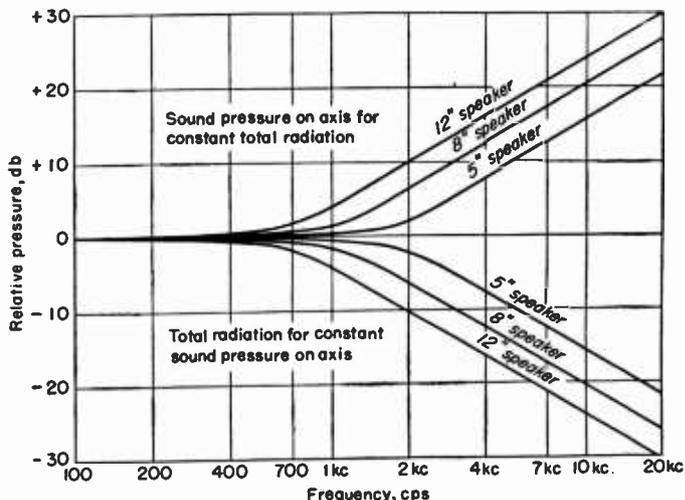


FIG. 17. Effect of concentration of radiation on piston axis at high frequencies for the case of pistons radiating constant total power (top curves) and the case of constant pressure response on the axis (lower curves). Because of cone flexing the concentration is less marked in actual diaphragms where the equivalent piston diameter at high frequencies approaches about six-tenths the actual cone diameter.

response must rise appreciably at high frequencies. The variation in relative response with angle up to the angle for which the first minimum occurs is shown in Fig. 18. The response on the axis has been arbitrarily adjusted to the same reference level in all curves. At high frequencies the effective area of an actual cone is reduced by flexing, so that the directivity of actual cones is somewhat less than that shown for the piston.

Typical directional curves for 6- and 10-in. (designating size) speakers are shown in Fig. 19. The axial response is assumed equalized to give flat response and the relative response for other angles is shown. Typical directional curves for a 6- by 9-in. (designating size) elliptical speaker are also shown. These show that

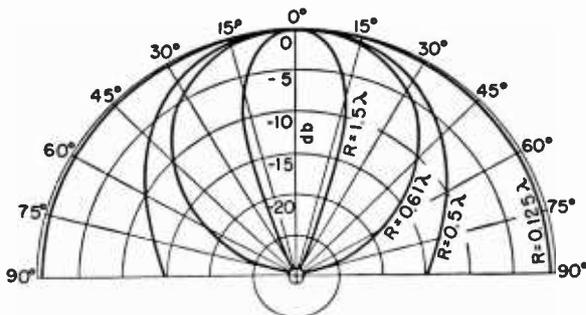


FIG. 18. Variation in relative response with angle up to the angle for which the first minimum occurs. The response on the axis has been arbitrarily adjusted to the same reference level in all curves.

the directional response of this elliptical speaker in the plane of the *minor* or short axis is roughly comparable to that of a circular speaker with a diameter equal to the minor axis. The directional response in the plane of the major or long axis is worse than that of a 10-in. circular speaker up to about 6,000 cycles. Above this frequency it is

better. Contrary to popular belief the elliptical speaker should be mounted with its major axis *vertical* to get broadest distribution in the *horizontal* plane. This corresponds to the theoretical predictions of Stenzel. This same effect is present in rectangular mouth-shaped horns where the middle and middle h-f response is *most directional* in the plane of the *broadest* horn mouth dimension unless partitions, or separate

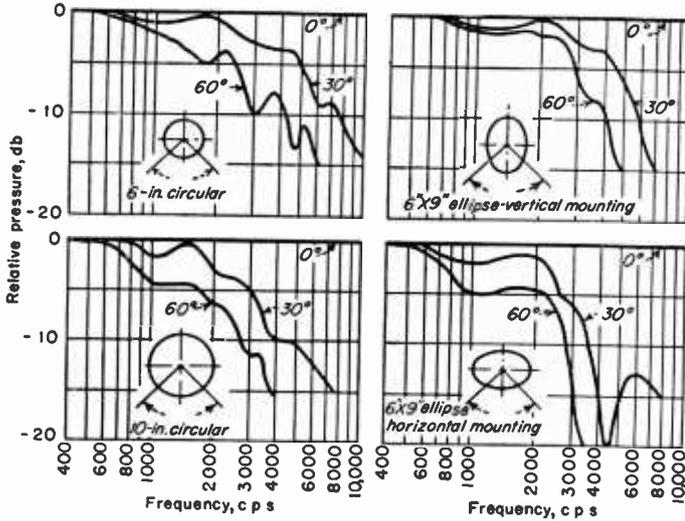


FIG. 19. Experimentally determined directional characteristics for two circular and one elliptical diaphragms. The elliptical diaphragm has the *broadest* characteristic in the horizontal plane when its major (long) axis is *vertical*.

cells are used. Even then the effect is present when the wavelength is comparable to the smaller dimension of the entire mouth.

9. Horns. A horn is a tapered acoustical transmission line used to couple the impedance it sees, looking back into the diaphragm, as efficiently as possible to the

load it sees looking out of its mouth (see Motors). The small end of a horn is called its *throat*, the large end its *mouth*. If its mouth has an infinite flange or baffle, the radiation impedance it sees is approximately the impedance given in Fig. 11. If there is no flange, the radiation resistance is half and the reactance approximately seven-tenths this value at low frequencies. At high frequencies the flange does not alter the impedance.

It has been found that many of the horns commonly used in loud-speakers belong to the hyperbolic exponential or "catenoidal" horn family, in which the

FIG. 20. Longitudinal section of straight-axis circular cross-sectional horns of hyperbolic exponential or catenoidal family.

cross-sectional area, A , is related to the axial distance, by

$$A = A_t \left[\cosh \left(\frac{x}{x_0} \right) + T \sinh \left(\frac{x}{x_0} \right) \right]^2 \tag{6}$$

where A_t is the throat area, x_0 is a constant fixing the axial scale of length, T is a constant determining a member of the general family, and the cosh and sinh are the

hyperbolic exponential cosine and sine functions, respectively. The longitudinal sections of these horns for various values of T are shown in Fig. 20 for straight-axis circular horns.

When T is infinite, a conical horn is obtained. For T equal to unity the bracketed expression reduces to e^{2x/x_0} yielding the equation for the familiar exponential horn. At T equal to zero the shape is that of the hyperbolic cosine, with zero slope at the origin.

The performance of a horn depends principally on the throat impedance and its dependence on frequency. While waves reflected from the mouth introduce fluctuations into the throat impedance, it has been found that the average impedance is closely that of a horn with no reflected wave. The throat impedance in mechanical ohms of catenoidal horns with rigid, nonabsorbent walls and negligible reflected waves is given by

$$Z_t = R_t + jx_t$$

$$= A_t \rho c \frac{[1 - (f_c/f)^2]^{1/2} + j(Tf_c/f)}{1 - (1 - T^2)(f_c/f)^2} \quad (7)$$

Where ρ , c , and T have been defined, f is a frequency, and f_c is the cutoff frequency, given by $f_c = c/2\pi x_0$. Thus the reference axial length x_0 is of fundamental importance in determining the behavior of the impedance. In Fig. 21 is shown the behavior of R_t for various values of T . The product ρc has been replaced by 41.4, its value in cgs units at room temperature.

For frequencies below f_c the resistance R_t becomes zero, and the horn cannot lose energy to the medium. Of course, in practice there is always enough reflected wave near the cutoff frequency to permit some radiation below this, so that the transition

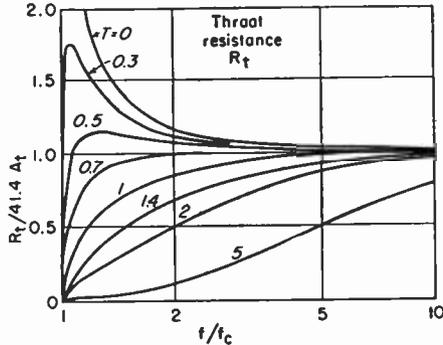


FIG. 21. Frequency dependence of throat resistance for horns of various values of T . Below f_c the resistance is zero.

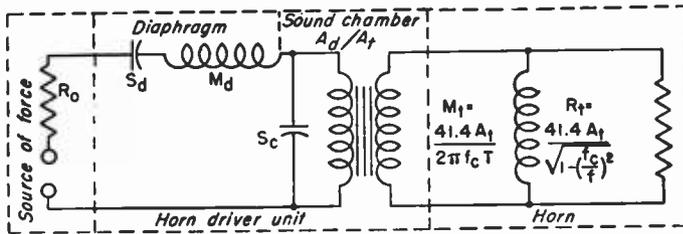


FIG. 22. Equivalent electrical circuit of mechanical elements of horn speaker. The generator impedance R_0 increases with the efficiency of the speaker.

is not too abrupt. In many designs the cutoff of the system comprising amplifier, horn-driver unit, and horn may be judged aurally to be 10 or 20 per cent above that of the horn alone, often due to impedance effects arising from other elements of the system.

The behavior of the horn throat impedance is conveniently expressed by an equivalent electrical circuit of the mechanical elements, as in Fig. 22. The constant mass, m_s , depends inversely on T and is shunted by the frequency dependent resistance, R_t . These two elements account for the high-pass characteristic of horns. The driver unit is coupled to the horn by the sound chamber, which is the volume, V_s , terminated by the horn-throat area, A_t , and by the diaphragm area, A_d . This acoustic element may be represented by S_c , the shunt stiffness of the enclosed air plus an ideal trans-

former of turns ratio A_d/A_t . Referring again to Fig. 22 the diaphragm mass and stiffness, m_d and s_d , are driven by a source of force with internal impedance R_0 mechanical ohms. It is this "generator resistance" to which the horn must be properly coupled to attain a speaker of high efficiency.

The reactive elements in the equivalent circuit can be chosen on the basis of filter theory, but the response so obtained may not be the type desired. In general, the ratio A_d/A_t may be chosen so as to produce approximate equality of source and load resistances in the "pass" or transmission band. In this case the resonant frequencies of the diaphragm alone, and of the stiffness s_c with the mass $(A_d/A_t)^2 m_t$, are placed so as to emphasize the desired frequency regions.

An important horn design parameter is the cutoff frequency, which must be properly related to the resonant frequency of the diaphragm to produce the desired l-f response.

FIG. 23. Cross section of annular V-shaped diaphragm and sound chamber used to reduce destructive interference in sound chamber at high frequencies.

Once a motor and diaphragm have been selected, the horn throat area is usually fixed by the resistance-matching consideration mentioned, and T is chosen from Fig. 21. Thus the desired horn contour may be plotted from Eq. (7), and continued until the mouth is sufficiently large to minimize reflection. This occurs at a mouth diameter greater than about $\lambda/3$ at the low cutoff frequency. An equivalent statement is that the product of the cutoff frequency and the mouth diameter (in inches) be greater than 4,500. This product can be considerably reduced if the speaker is operated under matched resistance conditions, for then a 4:1 change in load resistance results in but 2 db reduction in output. Thus the irregularities due to wave reflections at both ends of the horn do not show up strongly in the radiated energy.

The dimensions of the sound chamber are fixed by considerations of distortion and h-f cutoff. If the motion of the diaphragm is comparable to the clearances, then the stiffness, s_c , will vary throughout the cycle and lead to intermodulation distortion. However, if the clearance is increased, the value of s_c is lowered, hence its reciprocal the capacitance raised, thus shunting the horn load excessively at high frequencies. To calculate the amplitude of diaphragm motion there may be used the expression for the radiated power

$$P_a = 10^{-7} (\omega y_d)^2 A_d^2 \frac{R_t}{A_t} \quad (8)$$

evaluated about an octave above the cutoff frequency. Here P_a is the acoustic power in watts, and y_d the diaphragm amplitude in centimeters. Note that frequency enters into both ω and R_t .

At high frequencies such that the wavelength is comparable to the diaphragm

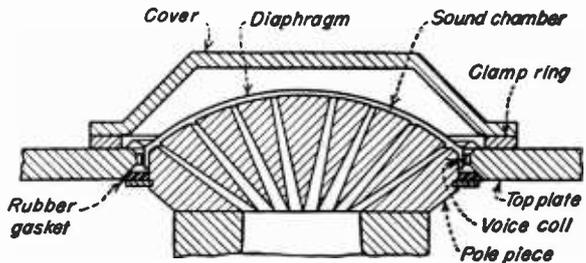
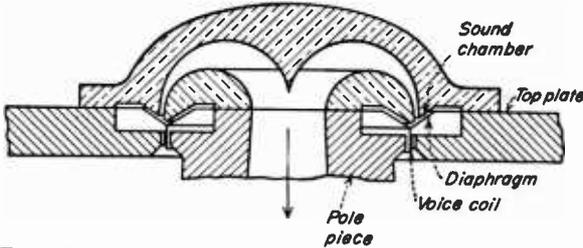


FIG. 24. Cross section of a diaphragm and sound chamber in which annular slots are used to reduce destructive sound-chamber interference at high frequencies.

diameter, transverse waves in the sound chamber produce cancellation effects which sharply reduce the radiated power. This condition can be ameliorated by subdivision of the throat into properly placed annular slots, as in Fig. 23, or by an annular diaphragm, as in Fig. 24. In these structures the sound chamber stiffness, s_c , may be considered a lumped constant equal to $1.37 \times 10^6 A_d/V_c$ dynes per cm over most of the frequency range, where A_d and V_c are defined above. The capacitance is the reciprocal of this value.

In Fig. 23 is shown a design chart for exponential horns. The areas are taken perpendicular to the direction of travel of the sound wave in the horn, and thus apply to folded and reentrant horns.

With a horn the directionality of the speaker may be controlled almost independently of the response, because the directional properties are fixed mainly by the mouth size and shape. Thus it is possible to obtain in horn speakers more control over response, efficiency, and directivity than in direct-radiator types.

In Table 3 are listed the absorption coefficients for some representative proprietary acoustical materials and objects, and the total absorption for persons and seats, for use in Eq. (12).

DIAPHRAGMS

10. Principle of Operation. The diaphragm is the part of the speaker which couples the radiation impedance to the speaker motor impedance. In the usual hornless or direct-radiator speaker, the force exerted by the motor is localized, and this must be transmitted to the acoustic load which is spread over a larger area. To do this effectively and to add as little as possible to the impedance, the diaphragm is made as rigid and light as possible. Conical shaped diaphragms are commonly used because they are rigid. The term "cone" is frequently used as a synonym for diaphragm in direct-radiator speakers.

The usual conical diaphragm may be thought of as a continuous mechanical transmission line radiating acoustic energy from each element of area. There exist both radial waves which travel from the driven region to the edge and are reflected, and circumferential waves which travel around the diaphragm. Both occur in various combinations depending on the mode of vibration and the symmetry of the driving force, diaphragm, and supporting members. Radial waves which travel from the driving point to the edge are reflected, and circumferential waves which travel around the cone, both occur in various combinations depending on the "mode of vibration."

The lowest frequency mode and the simplest one is the one in which the effective radial wavelength of the diaphragm, including the edge termination, is $\lambda/4$. (This must not be confused with a $\lambda/4$ in air at the same frequency.) At this frequency, which ranges from 700 cycles in large to 1,400 cycles or more in small conical diaphragms, no circumferential wave is present, and all parts of the cone move in phase. The displacement is a maximum at the apex and a minimum at the flexible annulus which supports the outer edge and terminates the transmission line. The impedance of this termination plays an important part in the diaphragm behavior, especially at frequencies near the fundamental resonance of the diaphragm and motor and in the 1,000- to 2,000-cycle range. At the l-f resonant frequency it may be considered a lumped stiffness whereas at the h-f end its distributed mass and stiffness are important.

At frequencies below the lowest mode of the cone itself all parts of the cone move in phase, and the cone behaves approximately as a piston unless the annulus stiffness increases rapidly with displacement, in which case the cone may flex at even very low frequencies. The annulus is frequently made this way deliberately in inexpensive speakers to produce l-f distortion and substantially increase their loudness by radiating most of the energy at harmonic frequencies. Unfortunately, intermodulation of low and high frequencies then also occurs, which makes the high end sound rough or garbled when a strong low note is reproduced.

11. Size. It has been found experimentally that at low frequencies the effective area of the cone is its projected or base area. This is approximately the "cone" size

where this is defined for a circular cone as "the diameter to the nearest $\frac{1}{4}$ in. of the minimum circle determined by the tangency of the cone and a plane touching its base."¹ This is not to be confused with the *designating* size of a loud-speaker which is commonly used in describing a speaker.

The designating size of a loud-speaker employing a circular radiator shall be twice the maximum radial dimension, measured to the nearest eighth-inch, of the front of the speaker except that the designating size shall not exceed the maximum diameter of the unsupported portion of the vibrating system by more than 25 per cent.²

This definition is intended to limit the amount of functionally useless cone housing included in the designating size. Representative cone sizes for various speaker-designating sizes are shown in Fig. 12.

In direct radiator speakers and at low frequencies the radiation resistance is proportional to the fourth power of the radius (square of the area) and the reactance to the cube of the radius. The resistance-reactance ratio or power factor of the radiation impedance is there-

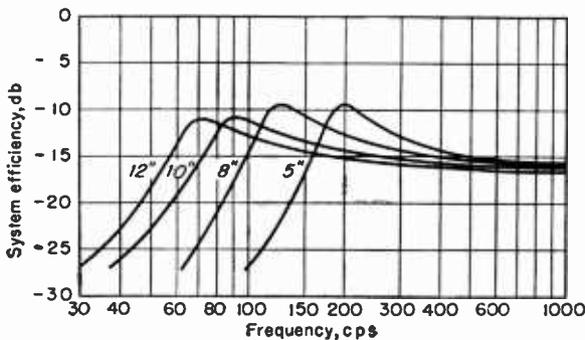


FIG. 25. System efficiency of four speakers using identical moving-coil motors but different cone sizes (calculated). Designating size of speaker is given.

fore proportional to the piston radius. For constant radiated power the piston displacement varies inversely with area. With fixed amplitude the radiated power is proportional to the square of the area at a given frequency, or equal power may be radiated at half the frequency if the area is increased by a factor of 4. The upper limit to diaphragm size is set by the increased weight *per unit area* required to get a sufficiently rigid struc-

ture. The nature of the acoustic load (horn, enclosure, cabinet, etc.), space limitations, cost, and the motor employed also control size. It is customary to increase the size of the motor as the diaphragm size is increased, since the mechanical impedance looking back into the motor (voice coil, say) should go up as the impedance looking into the driving point of the cone rises to maintain good energy transfer. If a large cone is put on a small motor, the displacement and distortion for a given acoustic output drop and lower frequencies may be reproduced with the same distortion, but the efficiency in the mid-range may actually drop. These effects are illustrated in Fig. 25, in which the calculated system efficiency of four speakers using different size cones but the same motor are shown. Minimum cone weights, found to be satisfactory experimentally, and average mechanical resistance and resonant frequencies were assumed. The motor is an intermediate size normally employed on 8-in. speakers but frequently used on all four diaphragm sizes. Speaker efficiency, even at low frequencies, is therefore not limited by cone size. The cone size must be large, however, if appreciable power is to be radiated with reasonable cone excursions at low frequencies.

12. Shape. The most efficient shape at low frequencies is circular. This is also the most satisfactory structurally. Theoretical and experimental investigations have shown that an ellipse with a major to minor axis ratio of two, and a two-to-one rectangle have an average of 5 and 7 per cent lower radiation resistance in the useful l-f range

¹ Radio Manufacturers Association, definition M5-111.

² Radio Manufacturers Association, definition M5-110.

than a circle of the same area. The loss is progressively greater as the shape departs still further from circular. In spite of the appeal of elliptical and other diaphragm shapes, which were used in early magnetic speakers and even in more recent European moving-coil speakers to a limited extent, before they were introduced in this country, their disadvantages have prevented their general adoption.

At high frequencies all pistons have the same radiation resistance per unit area, but most cones cannot be considered pistons, both because they are not flat and because their radial length exceeds $\pi/4$ and flexing is therefore important.

The shape of the cross section or profile of the diaphragm depends on the application and response desired. Straight side cones are usually employed when good 2,000- to 5,000-cycle response is required and when reproduction above 6,000 to 7,000 cycles may actually be undesirable. This is frequently the requirement of public-address and phonograph systems where noise and distortion are otherwise objectionable. Curved cones improve the response above 6,000 to 7,000 cycles by providing a diaphragm impedance, viewed from the voice coil, which has a more uniformly high negative reactance and therefore absorbs more power from the high positive reactance (due to the voice-coil mass) seen looking back into the voice coil. This improvement is obtained at the expense of 2,000- to 5,000-cycle response and with a weaker cone structure, with the result that straight cones predominate by a factor of 10 or more in actual use.

13. Material. Hard, impregnated or filled, and pressed or calendered papers are used when loudness efficiency and apparent h-f response are important. Radiation resistance provides very little dissipation in direct radiator cones; hence, by using a paper having low internal flexural losses, the conical transmission line is made to have strong resonances. Nearly all direct radiator speakers now use material of this type. The transient response of diaphragms of this type is necessarily poor since noncenter moving modes of the cone are inappreciably damped by the motor. Soft, loosely packed, or felted blotterlike cones are used when some loss in h-f response can be tolerated and a smoother response curve with reduced transient distortion is required. The loudness efficiency of high-loss cones of this type is several decibels lower than that of low-loss cones.

Felt, leather, rubber, and similar materials have been used as the annulus to terminate the conical transmission line in a low-stiffness high-resistance material. Their effect is to add considerable dissipation to the cone at the termination, resulting in reduced reflection of the flexural wave. The effect is similar to that obtained in soft cone materials where, however, the dissipation is distributed along the line. These materials are now used infrequently because they increase the cost, lower the loudness efficiency, and necessitate larger air-gap clearances than are used when the annulus is made an integral part of the diaphragm.

14. Breakup Subharmonics. The term *cone breakup* is sometimes applied to the flexing or wave-transmission process in a cone. Since there is nothing discontinuous in the process to suggest the word "break" and to avoid ambiguity, it is suggested that this term be applied only in the other sense in which it is used, to name the process which results in the generation of subharmonics.

If the apex of a cone is driven with an adequate sinusoidal force at certain critical frequencies, the radiated wave contains not only the fundamental and integral multiples (harmonics) of it but also a frequency corresponding to half (and rarely to a quarter) that of the fundamental and integral multiples of this subharmonic. While distortion in the motor may contribute to this, some unpublished research has shown that the cone is the important source.

The half frequency appears very suddenly at a critical input as shown in Fig. 26. To simplify the graph, the fundamental and usual harmonics, which would include even multiples of the subharmonics, are not included. From the total (rms) harmonic curve H_{Σ} which includes these uncharted terms, we see that negligible rise in total distortion occurs when the subharmonic begins. The ear, however, reports a large increase because the pitch sense of the output has dropped an octave and the distortion

has a high annoyance or objectionability factor. This type of distortion is not so important as is frequently supposed, however, because (1) it occurs only in limited frequency regions; (2) it does not occur below a moderate, critical level; (3) the time required to start it is large unless the force is large; (4) the spectral composition of speech and music are such that the probability of its production is small. Because of the statistical improbability of its frequent occurrence, it would be uneconomical to design most systems to avoid completely this occasional distortion.

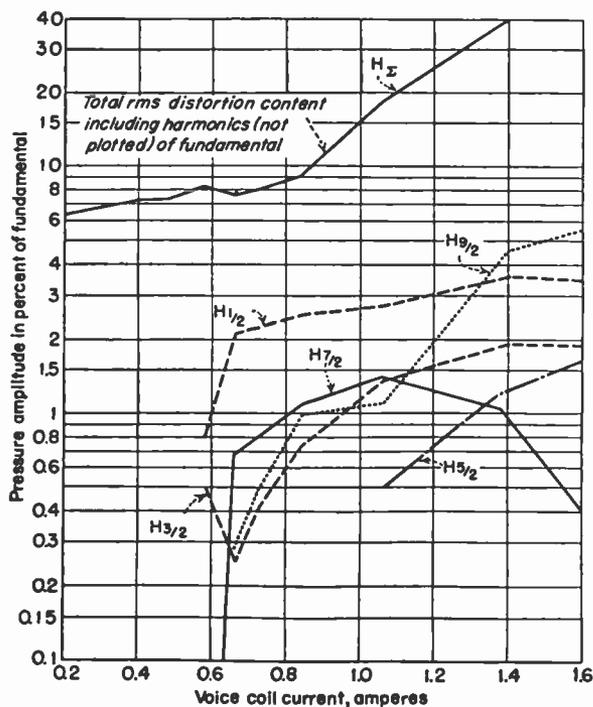


FIG. 26. Variation in radiated distortion products with voice-coil current showing abrupt start of half frequency (subharmonic) and odd multiples of it. Subscript indicates factor by which fundamental frequency is multiplied to obtain frequency of indicated distortion product.

MOTORS

A loud-speaker motor converts electrical into mechanical energy and couples the electrical signal source as efficiently as possible to the mechanical impedance seen looking into the diaphragm which it drives.

15. Force Factor. The mechanical circuit of a speaker motor experiences a force when a current is applied to the electrical terminals. The (complex) ratio of this force when the mechanical circuit is blocked (infinite impedance) to the current which produces it is the force factor. Since force is analogous to voltage (in the most commonly used system of analogues), the force factor is analogous to mutual impedance between two electrical circuits. It differs from the conventional electrical mutual impedance in that it makes no contribution to the electrical impedance when the mechanical circuit is blocked (secondary open-circuited) because its counter emf is due only to motion of the mechanical circuit, and in that the force factor in magnetic systems has opposite signs when viewed from the electrical and mechanical circuits.

In usual circuit notation $z_{12} = -z_{21}$ (not $z_{12} = z_{21}$ as in the electrical case). Since only the product of the force factors looking in both directions is involved in the following equations this will be called M^2 .

The normal impedance of a speaker is defined as the impedance measured or seen at its signal terminals when operating normally with its proper acoustic load. The

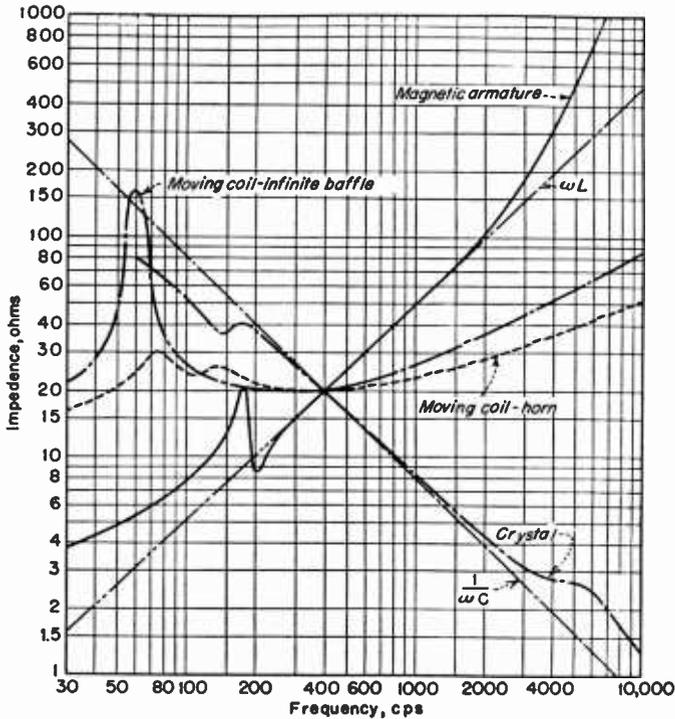


FIG. 27. Magnitude of normal input impedance of various speakers all adjusted to same value at 400 cycles to simplify comparison. Magnitude of reactance of pure capacitance and inductance shown for comparison.

normal impedance z_N of moving coil and magnetic armature speakers is

$$z_N = z_s + \frac{M^2}{z_m} \tag{9}$$

where z_s = blocked electrical impedance of the speaker

z_m = total mechanical impedance seen by the mechanical circuit including diaphragm and acoustic load

In moving-coil speakers $M^2 = B^2 l^2$, where B is the average radial flux density which the coil embraces and l is the conductor length. In balanced magnetic armature speakers $M^2 = 4B_0^2 N^2 / R^2$, where B_0 is the steady flux density in the gaps, N is the number of turns on the voice (armature) coil, and R is the effective reluctance of the alternating flux path (see Magnetic Armature).

A two-terminal load impedance absorbs maximum power from a two-terminal source when the impedance of the load is the conjugate of the impedance measured or "seen" at the source terminals. The conjugate impedance is one having the same resistive or real part and a reactive or imaginary part equal in magnitude but opposite in sign.

This holds for acoustical and mechanical circuits as well, but in these the terminals are not always so readily determined.

The speaker motor therefore absorbs maximum energy from the source, regardless of the complexity of the source network, when its normal impedance is the conjugate of the source impedance. The usual source is a vacuum tube, and its associated

loud-speaker should ideally have a normal impedance which is a constant resistance. From Fig. 27 we see that this requirement is most closely met by moving coil or "dynamic" speakers.

16. Magnetic Motors. By the IRE definition, "A magnetic speaker is a loud speaker in which the mechanical forces result from magnetic reactions." This includes both moving-conductor or moving-coil (dynamic) and magnetic-armature speakers.

17. Moving Coil. A moving-coil motor is one in which the mechanical forces result from magnetic reactions between the field of the moving coil and the applied steady radial field in the air gap. A section of half of a moving-coil speaker is shown in Fig. 28.

Moving-coil motors are now used almost exclusively because (1) their electrical impedance permits good energy transfer from the source, (2) the large amplitudes required by the popular direct radiator diaphragms are obtained conveniently with minimum nonlinear distortion, (3) the mechanical impedance of the moving element may be made low, (4) the structure is simple and rugged mechanically, and (5) the cost is low.

FIG. 28. Sectional view of small moving-coil speaker showing structural simplicity.

The impedance seen at its electrical terminals when the coil is rigidly blocked is called the *blocked impedance* of the motor or speaker. This is approximately a high

resistance and low inductance in series (R_e and L_e in Fig. 29) and is therefore easily coupled to a vacuum tube. Near the fundamental resonance of the speaker the impedance rises, and, if a low-impedance source is used, the mismatch reduces the energy absorbed.

Moving-coil speakers are sometimes called *electrodynamic* or briefly *dynamic* speakers. Both terms have been applied for many years to speakers having either

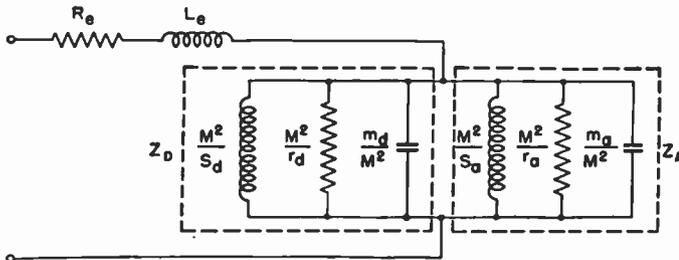
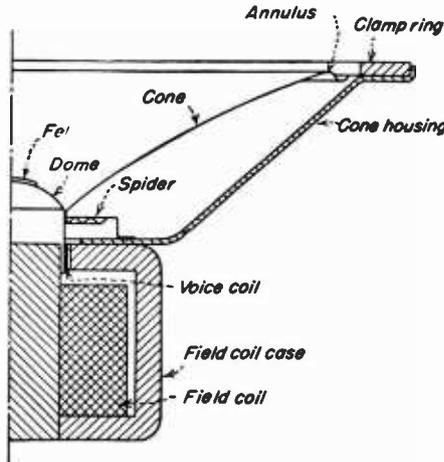


FIG. 29. Equivalent 1-f electrical circuit of moving-coil or magnetic armature speaker in a total enclosure or in an infinite baffle. In the latter case the enclosure stiffness S_a is zero and its equivalent electrical inductance infinite.

resistance and low inductance in series (R_e and L_e in Fig. 29) and is therefore easily coupled to a vacuum tube. Near the fundamental resonance of the speaker the impedance rises, and, if a low-impedance source is used, the mismatch reduces the energy absorbed.

Moving-coil speakers are sometimes called *electrodynamic* or briefly *dynamic* speakers. Both terms have been applied for many years to speakers having either

electromagnet (or "energized") or permanent magnet fields. The prefix "electro" in electrodynamic has nothing to do with the source of steady flux in the gap.

18. Permanent and Electromagnets. Magnetic speakers require a source of magnetomotive force to provide the steady flux. If the current source is hum-free and therefore the flux absolutely steady, the voice coil cannot distinguish between a given flux density due to permanent and electromagnets. The efficiency of any electromagnet speaker can be equaled or excelled by a permanent magnet if cost is neglected. In small motors the permanent magnet type is usually less expensive. In inter-

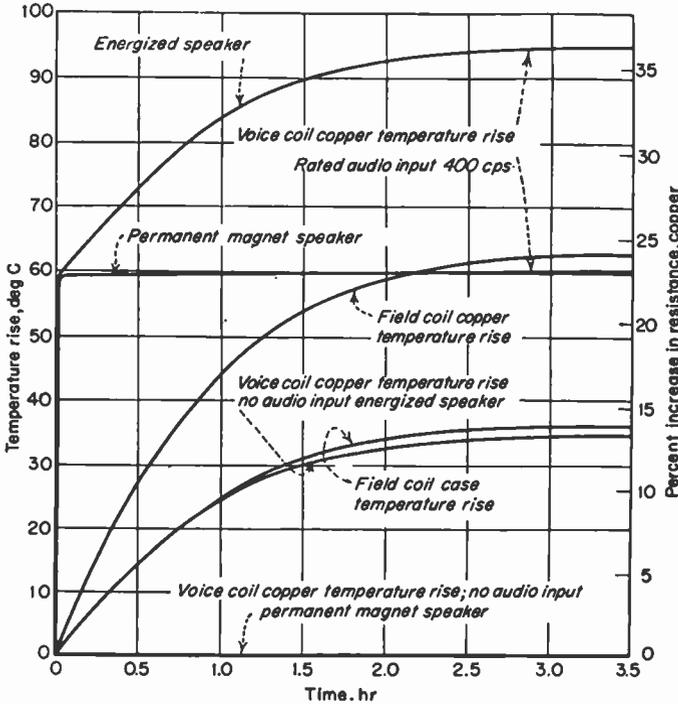


FIG. 30. Reduced voice-coil temperature rise in permanent-magnet speakers compared to energized types. Abnormal voice-coil temperature rise when rated complex-wave input is applied at 400 cycles also shown.

mediate-size motors the cost of the two types is comparable if the cost of a source of field power is included. In larger motors the permanent magnet type is more expensive. In extensive speaker systems the installed cost of permanent magnet types is frequently lower because of simplified low-cost wiring.

The trend is toward the use of permanent-magnet speakers particularly when a special field current supply must otherwise be provided. The temperature of the electromagnet and consequently that of the voice coil rises with time as shown in Fig. 30. The field coil resistance rises, lowering the field current and flux density. The higher voice-coil impedance and reduced flux reduce the speaker efficiency. The higher voice-coil temperature reduces the permissible signal input power in voice-coil temperature-limited speakers.

The temperature rise when the rated complex-wave input is applied at a single frequency (400 cycles) in a typical intermediate-size radio speaker is also shown. The single-frequency rating is normally much less than the "complex-wave" (speech

and music) rating since in the latter case advantage is taken of the high ratio of peak to average power (see Secs. 1, 2, 35).

19. Magnetic Armature. "A magnetic armature speaker (or motor) is a magnetic speaker (or motor) whose operation involves the vibration of the ferromagnetic circuit." The shorter term "magnetic" may be used where no confusion will result with moving-conductor or moving-coil speakers, which are also by definition magnetic speakers. A cross-sectional view of a balanced armature motor of this type is shown in Fig. 31. Flux increases in one pair of pole faces and decreases in the other pair, when current flows through the voice coil and when the armature moves, resulting in operation analogous to a push-pull tube circuit. The voice coil does not move and therefore is made relatively large. The resulting high inductance plus distributed capacitance in high impedance types accounts for the large rise in impedance at high frequencies (see Fig. 27). This makes it difficult to couple it to a tube properly. To get high efficiency the armature pole-piece clearance must be small, and this leads to instability of the armature and a limitation on its displacement. These factors

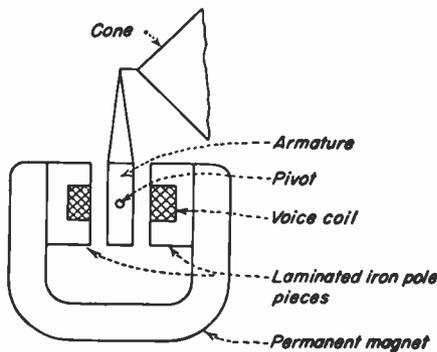


FIG. 31. Sectional view of balanced magnetic-armature speaker.

The electrode clearance must be small or the steady polarizing potential which is applied must be large to get high efficiency. The former limits the diaphragm amplitude, and the latter causes rapid disintegration of any flexible dielectric used to support the electrode.

21. Crystal Speaker. "A crystal speaker (or motor) is a speaker (or motor) in which the mechanical forces result from the deformation of a crystal having converse piezoelectric properties." The crystal has a high mechanical impedance viewed from the driving point. Only a small displacement is possible without distortion or crystal fracture, so a mechanical transformer or lever arm is used when moderate excursions are required. This leads to mechanical complications, particularly at high frequencies where the transformer is not ideal, and to added cost. The application of this type has therefore been largely limited to h-f speakers in which the diaphragm amplitude is small. The blocked impedance is that of a leaky capacitance. The normal impedance of an 8-in. unit is given in Fig. 27.

COMPLETE LOUD-SPEAKERS

The more important characteristics of a complete speaker system, which includes an electrical source of known impedance, motor, diaphragm, and known acoustic load, are its efficiency-frequency (including response-frequency and impedance), directional, and distortion characteristics (see Tests).

22. Efficiency-frequency Characteristic. The energy efficiency, or simply efficiency, of a loud-speaker is the ratio of the useful acoustic-energy output to the signal-energy input. The "absolute" or system efficiency is the ratio of the useful acoustic

plus mechanical difficulties in construction and maintenance have reduced the acceptance of the magnetic-armature type.

20. Condenser Speaker. "A condenser speaker (or motor) is a speaker (or motor) in which the mechanical forces result from electrostatic reactions." They are really large capacitors in which one flexible electrode is free to move and to act as a diaphragm. In push-pull types the flexible electrode is mounted between two perforated fixed electrodes resulting in cancellation of the even harmonics which occur in the two-plate type. Its blocked impedance is that of a capacitance, and it is therefore difficult to couple to a vacuum tube.

energy output to the signal energy an ideal load would absorb from the signal source. The latter definition is a practical one in that it penalizes the speaker for its inability to absorb maximum power from the source. At a resonant frequency of a speaker the two efficiencies frequently differ by a factor of 10 or more.

If the effective internal resistance of the source and its ideal resistance load (both seen from the voice coil) are r_s , then the absolute efficiency is given by

$$\text{Absolute efficiency} = \left| \frac{4r_s M^2 r_r}{[z_s + (M^2/z_m)]^2 z_m^2} \right| \quad (10)$$

where z_s = blocked voice-coil impedance plus r_s ,

z_m = total mechanical impedance of the mechanical mesh including diaphragm radiation and air load

r_r = total radiation resistance seen by diaphragm

M is defined under Force Factor. The vertical lines indicate that the absolute value is to be taken.

In a direct-radiator moving-coil loud-speaker mounted in an infinite baffle the system efficiency is not constant with frequency. However, there is one region, the piston range, in which a fortuitous combination of characteristics yields a reasonably constant efficiency. For the ordinary large direct-radiator speaker this usually occurs over a substantial portion of the 100- to 800-cycle range. In smaller sizes this range may be limited to 400 to 1,000 cycles. Here most cones vibrate as rigid pistons, so that their radiation is closely that of a piston of the same projected area. Next, $2\pi R/\lambda$ is less than unity, so that the radiation resistance r_r varies with the square of the frequency (see Fig. 11). The frequency at which the motor is resonant usually lies well below that at which $2\pi R/\lambda$ is unity, and since the diaphragm is pistonlike, the mechanical reactance in the piston range is predominantly masslike. Also, the impedance of the voice coil is largely resistive and nearly constant. If it is coupled to a source of equal resistance and if radiation from only one side of the diaphragm is considered, it turns out that the piston-range system efficiency is equal to about $5 \times 10^{-11} \alpha E d^4 / m^2$. Here α is the ratio of the volume of the voice coil conductor to the total gap volume, E is the magnetic energy in the gap, in ergs, d is the effective diameter of the diaphragm, in centimeters, and m is the sum of the mechanical and fluid masses of the moving system, in grams. From this expression it is seen that the contribution of the magnetic structure to the speaker efficiency is in terms of the gap energy, rather than the flux density or the total flux in the gap. Since the gap energy is roughly proportional to the magnet weight or the field-coil copper weight, it is seen that economic factors play a large part in determining the efficiency, at least in the piston range.

The 400-cycle system efficiency of the speakers commonly used in radio receivers ranges from 1 to 4 per cent. The corresponding efficiency of direct-radiator speakers with very large motors ranges from 10 to 30 per cent. Efficiencies of this order are more readily obtained in horn speakers, but 30 per cent is rarely exceeded over an extended frequency range. Although higher values are frequently claimed, these values, if based on any measurements, are usually based on motional impedance measurements in which all horn, diaphragm, air, eddy-current, and hysteresis losses have been assumed to be useful acoustic radiation.

23. Response-frequency Characteristic. If a loud-speaker is to be used indoors, a graph showing the efficiency-frequency characteristic is probably the most useful single curve. If a loud-speaker is to be used outdoors, then we are primarily interested in its pressure response-frequency characteristic (see Tests).

24. Baffles, Enclosures, and Cabinets. "A baffle is a partition which may be used with an acoustic radiator to increase the effective length of the acoustic transmission path between front and back of the radiator." This term is usually reserved for a relatively flat baffle in which both sides of the diaphragm look into substantially a hemisphere (solid angle of 2π steradians). The term *directional baffle* is sometimes applied when one side of the diaphragm looks into a smaller solid angle. The baffle

then begins to take on the properties of a horn. There is no sharp line of demarcation, but there appears to be little reason for calling any structure which restricts the solid angle to less than $\pi/2$ (an octant of a sphere) anything but a horn.

If a baffle is used outdoors, appreciable destructive interference or pressure cancellation between the front and back waves of the speaker may occur at the listener's position at some frequency above the cutoff frequency. The frequency at which this occurs depends on the baffle size and listener location. Destructive interference at



FIG. 32. Irregular baffle shapes used outdoors to broaden frequency band of destructive interference between speaker front and back waves at listener's position.

the cone itself is usually unimportant except near the cutoff frequency of the baffle. To distribute this effect and make it cover a broad band, baffles shaped as shown in Fig. 32 are sometimes used. Since the effect depends primarily on the listener's location, no such simple result occurs indoors and a space average of the pressure in a moderate-size listening room shows no

such effect. Conventional rectangular baffles may therefore be used indoors unless the room approaches free-field or outdoor characteristics.

The equivalent l-f electrical circuit of a moving-coil or magnetic-armature speaker in an infinite baffle is shown in Fig. 29. Here R_e and L_e are the blocked voice-coil resistance and inductance. Z_D is the electrical equivalent of the diaphragm less air load. Z_A is the equivalent of the air load, except that in this case there is no stiffness S_a provided by the air load, so its equivalent inductance M^2/S_a is infinite. M is defined under force factor S_d , m_d and r_d are the effective diaphragm stiffness, mass, and resistance, respectively, and r_a and m_a are the radiation resistance and mass which may be determined from Fig. 11. Note that both sides of the diaphragm have radiation resistance and mass in this case and the values *per unit area* given in Fig. 11 must be multiplied by twice the piston area to give the r_a and m_a used in Fig. 29.

The magnitude of the impedance of a moving-coil speaker in an infinite baffle is shown in Fig. 27. The antiresonant impedance of the parallel circuit corresponds to the resonant frequency of the diaphragm and air load and is limited by the parallel value of the two resistances. These resistances are proportional to the square of the flux density and inversely proportional to the diaphragm and air (radiation plus dissipation) resistances. A high resonant impedance is therefore not necessarily undesirable, as is generally supposed, since it may be due to a high flux density and therefore mean a high efficiency over a wide frequency band.

The effect of the source impedance, connected to the input terminals, on response may be noted from this circuit. If the source resistance r_s is low, the speaker will absorb very little power at resonance and the acoustic output may not rise appreciably. The voice-coil resistance R_e and the source resistance r_s in series are effectively across the antiresonant circuit at low frequencies since the reactance of L_e can be neglected.

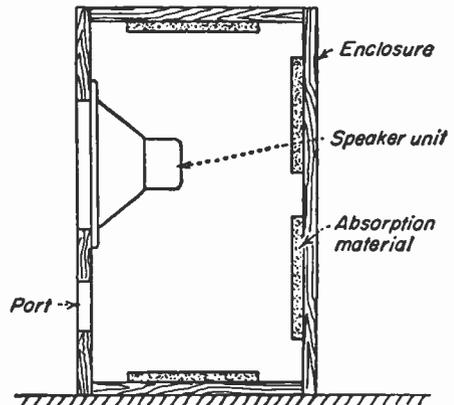


FIG. 33. One type of vented enclosure in which port area is large and placed near diaphragm to obtain maximum aid from mutual radiation resistance between diaphragm and port. Phase shift of backside radiation obtained by choice of circuit constants.

When the flux density is high, the Q of the antiresonant circuit alone is high but the source and voice-coil resistances then provide substantial shunt resistance and "electromagnetic damping."

The effect of this on the speaker response to a pulse is shown in the experimentally determined curves of Fig. 47. The minor irregularities in these damped sinusoids are due to h-f modes of vibration of the diaphragm. As the source resistance is raised, more power is supplied the speaker at antiresonance and at high frequencies where the effect of the voice-coil inductance is important and less is supplied in the mid-frequency range. The steady-state

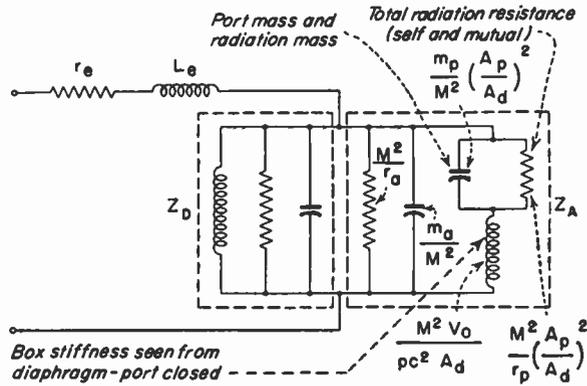


FIG. 34. Equivalent l-f electrical circuit of moving-coil or magnetic-armature speaker in vented enclosure.

response of a vented enclosure (see Figs. 33, 34, 35) as the source resistance or "impedance match" is changed is shown in Fig. 46. Here the response at 400

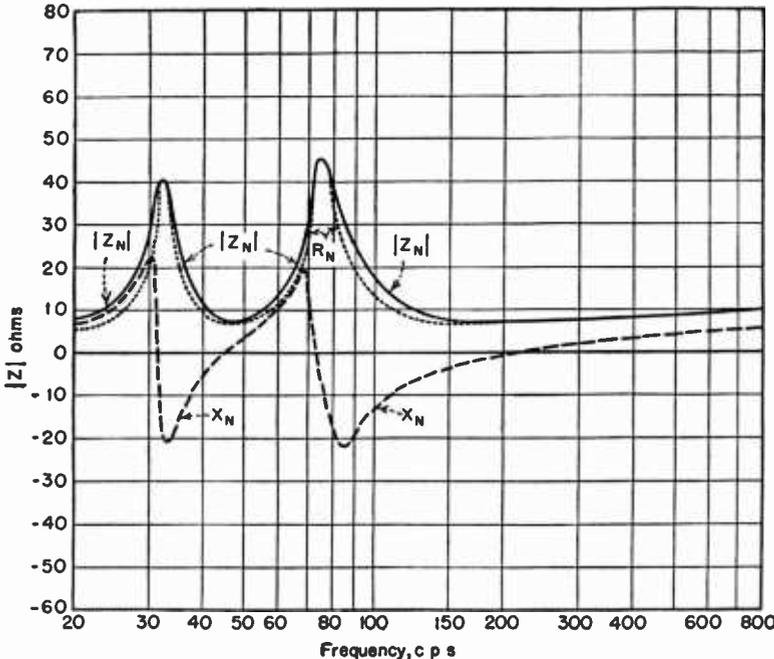


FIG. 35. Impedance of circuit of Fig. 34.

cycles has arbitrarily been adjusted to the same value as the source resistance was changed.

25. Total Enclosure. A total enclosure which prevents radiation from the back side of a diaphragm may be used to prevent destructive interference between the front and back waves from a diaphragm. This might be obtained by closing the vent or port in Fig. 33. When the wavelength exceeds four times the maximum enclosure dimension, the enclosure adds a *total* stiffness S_a viewed from the diaphragm $S_a = \rho c^2 A_d^2 / V_0$ cm per dyne, where A_d is the effective piston area of the cone and V_0 is the net volume of the box. The "capacitance" is the reciprocal of this value. This stiffness raises the natural frequency of the speaker. If the enclosure includes absorbing material, this stiffness will be altered by the reactance seen at the surface of the material. Each square centimeter will dissipate $P^2 \times 10^{-7} / r_B$ watt, where P is the sound pressure in the box and r_B is the resistance per unit area for sound of normal incidence on the absorbing material. From this the equivalent resistance in parallel with the box stiffness may be obtained.

If the volume is small enough or the natural frequency of the speaker out of the enclosure low enough, the enclosure and not the diaphragm stiffness will control the natural frequency.

The l-f equivalent electrical circuit of such an enclosure is shown in Fig. 29. Here z_D is the electrical equivalent of the diaphragm alone; S_d , r_d , and m_d are the stiffness, resistance, and mass of the diaphragm measured *in vacuo*. The electrical equivalent of the air load including radiation impedance is z_A ; S_a is the effective enclosure stiffness, r_a the total air or fluid resistance (enclosure dissipation if any, and radiation resistance), and m_a is the effective air (radiation-plus-enclosure) mass. Normally the parallel value of z_D and z_A or a single parallel "antiresonant" circuit is shown but the contributions of individual elements are then not as clear. Since the electrical circuit elements are *inverses* with respect to M^2 [see Eq. (9)], the stiffness appears as an inductance and the mass as a capacitance.

A total enclosure is sometimes called an *infinite* baffle. While it resembles one in preventing front and back wave interference, it has two important differences which make this designating term undesirable. The enclosure adds an air impedance to the rear of the diaphragm, which may be very different from that seen in an infinite baffle. An infinite baffle restricts the radiation to a hemisphere, and the radiation impedance seen by the diaphragm is given by Fig. 11. If the enclosure is used outdoors, the radiation resistance which the outside of the diaphragm sees at low frequencies is only half this value and the reactive part approximately seven-tenths this value. In practice the useful efficiency is almost halved at low frequencies. Indoors the impedance seen will depend on the environment as described in Radiation Impedance; also in Room Acoustics, below.

26. Vented Enclosures. The idea of putting a vent or "port" in an enclosure is very old. It was first done to provide "pressure relief." In more recent types, known as "vented" enclosures, detailed consideration has been given to the very important effect of the mutual impedance between the port and diaphragm. The port area is large and the port is near the diaphragm to increase the mutual radiation resistance and extend the frequency range over which it is effective (see Mutual Radiation Impedance). Such an enclosure is shown in Fig. 33. The effective or virtual diaphragm in the opening is coupled internally through the stiffness of the air in the enclosure to the diaphragm. It is also coupled externally. The equivalent l-f circuit is shown in Fig. 34. Impedance z_D corresponds to Fig. 29 and r_a and m_a correspond except that the mutual-radiation impedance must be added.

The vent and enclosure have therefore added one *LRC* circuit. The effect of this is to shift the backside-cone radiation by nearly 180 deg above the frequency at which the port mass m_p and box stiffness *viewed from the port* are resonant when the cone is blocked. For about one-third of an octave above and below this frequency most energy is radiated by the port. Although the diaphragm and port radiation are out of phase below this frequency, the port radiation greatly exceeds the diaphragm radiation near this frequency.

The enclosure is made as compact as possible. The port can be placed near the

diaphragm to increase the mutual-radiation resistance since the phase shift is not primarily due to transmission time delay but occurs because the acoustic circuit goes through antiresonance, the phase shift occurring suddenly at this frequency. In properly designed enclosures, advantage is taken of a large mutual-radiation resistance to improve the l-f efficiency. Very little absorption in the enclosure is wanted at low frequencies to take maximum advantage of backside radiation. At frequencies of several hundred cycles or more where the port radiates negligible sound the enclosure is made absorbent to avoid "box" resonance. The advantages of vented enclosures are (1) backside radiation is used to substantially increase the l-f output; (2) most of this output comes from the port which has no nonlinear diaphragm suspension stiffness to produce nonlinear distortion; (3) antiresonance of the enclosure occurs near the lower frequency of maximum radiation so the diaphragm amplitude is much less than it would be otherwise. The result of these factors on nonlinear distortion reduction is shown in Fig. 36 in which the effect of converting an open-back cabinet to a vented port enclosure of the same internal volume is shown. The change in response is shown in Fig. 37.

27. Transmission-line Speaker. The phase and amplitude of the backside radiation of a cone may be altered by coupling a conduit or acoustic transmission line to it. In early types the multiple-resonant properties of such a line were used to influence

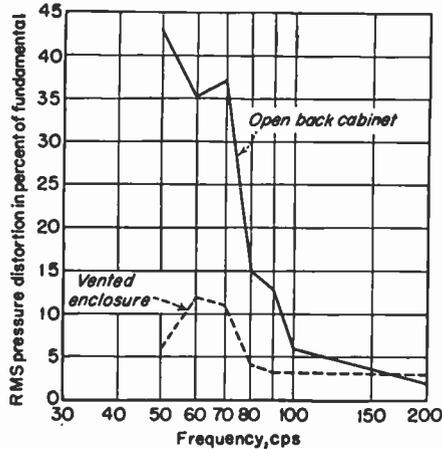


FIG. 36. Total distortion of open-back cabinet and of the same cabinet enclosed as shown in Fig. 33, same speaker and electrical input in both cases. L-f distortion much reduced because diaphragm sees high antiresonant impedance of enclosure and therefore has only small displacement whereas air in port (which lacks the nonlinear edge stiffness and nonuniform flux of the speaker) moves with large displacement.

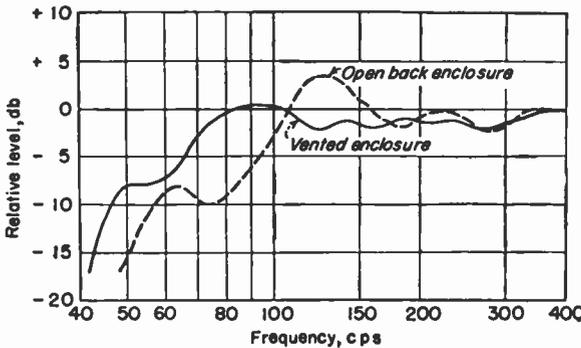


FIG. 37. Total sound power radiated by cabinet for which distortion is shown in Fig. 36. Shape of l-f response may be varied between wide limits depending on enclosure volume, port area, and speaker used.

the response. In a more recent type, known as the labyrinth, the line is folded to conserve space and made highly dissipative (see Fig. 38). Phase shift between the diaphragm and port or open end is due to time of transmission in the line. At very

low frequencies the line is a small fraction of a wavelength long, the phase shift is negligible, and the port and diaphragm radiation are out of phase. When the line is $\lambda/4$ long, it acts as an impedance inverter (as in the electrical case); the cone sees a high impedance, and the radiation from the port is a maximum. Nonlinear distortion is therefore reduced at and near this frequency. The resonant frequency of the diaphragm may be placed at this frequency to aid damping. Between this frequency and the one for which the line is $\lambda/2$ long, the port phase shifts gradually but maintains some component of its radiation in phase with the diaphragm (neglecting separation between the port and diaphragm) outside the line. Because of the infinite series of resonant and antiresonant frequencies of the line, high absorption must be introduced to prevent the production of objectionable resonances and radiated out-of-phase components of the port. Most of the rear-side radiation is therefore absorbed. The comparison of the response of an open-back cabinet and labyrinth is shown in Fig. 39.

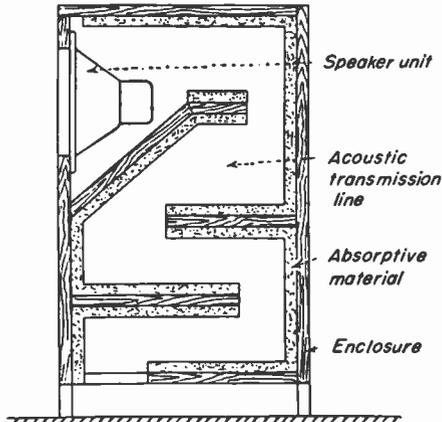


FIG. 38. Labyrinth type of transmission-line speaker. Phase shift of backside radiation is obtained by time of transmission delay in line.

the room, and the sound receiver or "sink," all as part of a unified dynamical system. This is required to bring out the interaction between source, sink, and room and their effects on the steady-state and transient aspects of sound transmission in the room.

In this theory the room is considered as an assemblage of resonators and the walls of the room as terminal impedances determining absorption and reflection. A rectangular room has a triple infinity of resonant frequencies. If the wall impedances are pure resistances, these frequencies are given by

$$f = 17,140 \left[\left(\frac{n_x}{l_x} \right)^2 + \left(\frac{n_y}{l_y} \right)^2 + \left(\frac{n_z}{l_z} \right)^2 \right]^{1/2} \quad (11)$$

where $n_x, n_y, n_z = 0, 1, 2, \dots$
 $l_x, l_y, l_z =$ dimensions of rectangular room, cm

The distribution of these "allowed" frequencies (at which resonance occurs) may be graphically shown as in Fig. 40 by a three-dimensional plot in "frequency space." Each vector to a lattice point is associated with a "natural frequency" or "normal mode" of the room. The shortest vector, corresponding to the lowest frequency, is determined by the longest dimension of the room. The direction of the vector from the origin to a lattice point indicates the direction of excitation of that frequency in the room, and the length of the vector is proportional to its frequency.

At low frequencies there may be an appreciable frequency interval between the

ROOM ACOUSTICS

28. Room Characteristics.

The trend in the theory of room acoustics is toward considering the source of sound, the room, and the sound receiver or "sink," all as part of a unified dynamical system. This is required to bring out the interaction between source, sink, and room and their effects on the steady-state and transient aspects of sound transmission in the room.

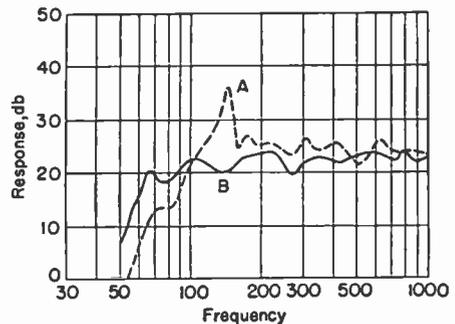


FIG. 39. Relative response of open-back cabinet A and labyrinth B. (After Olney.)

natural frequencies if the room is small. At high frequencies the number of natural frequencies in a given frequency interval is proportional to the square of the frequency.

29. Reverberation. Using this concept of multiple natural frequencies, the decay of sound in a room may be described as follows: Assume energy has been supplied the room until the energy level is constant; *i.e.*, the rate of absorption at the boundaries equals the rate of supply to the room. The resulting standing-wave system depends not only on the room and frequency but on the location and orientation of the source. When the source of energy is stopped, each individual mode of vibration of the room will decay exponentially, and the combined effect of these is called *reverberation*. Only the modes having allowed frequencies near the frequency of the steady-state excitation will contain appreciable energy.

By definition the reverberation time is the time required for the mean energy density in the room to drop 60 db. While this mean may be the result of a large number of rates of decay each of which is individually exponential, the combined value in general is not given by a single exponential term. This accounts for the fact that the slope of the mean-energy-density time-decay curves for the average room are not uniform and therefore for the fact that the apparent reverberation time depends on the time interval over which the decay is averaged. If the absorption is moderate the approximate reverberation time in seconds is given by

$$T = \frac{0.00161 V}{-A \ln (1 - \alpha)} \tag{12}$$

where V = room volume, cc

α = $(A_1\alpha_1 + A_2\alpha_2 + \dots)/A$ total room absorption

A_1, A_2 , etc. = areas, sq cm, having absorption coefficients α_1, α_2 , etc., respectively

A = total boundary area of room

Corresponding to this type of energy decay in the room, there is a growth curve. When a source suddenly emits energy, each of the excited modes absorbs energy in an exponential manner. This occurs until the asymptotic or steady-state value is reached after an infinite length of time. When this is reached, the acoustic power supplied the room must equal that absorbed at the room boundaries. By definition the absorption coefficient α of the boundary is the fraction of the incident energy absorbed for a specified angle of incidence. The intensity I of a sound wave [Eq. (2)] is a measure of the energy per square centimeter per second. The total intensity is made up of two parts: that due to the reverberant sound and that due to sound received directly from the source. The acoustic power radiated by the source is related to the total intensity by

$$P_o = \frac{I}{\frac{1 - \alpha}{A} + \frac{1}{4\pi r^2}} \quad \text{watt} \tag{13}$$

where P_o = radiated power, watts

I = total intensity, watts per sq cm

r = distance between source and point at which I is measured, cm

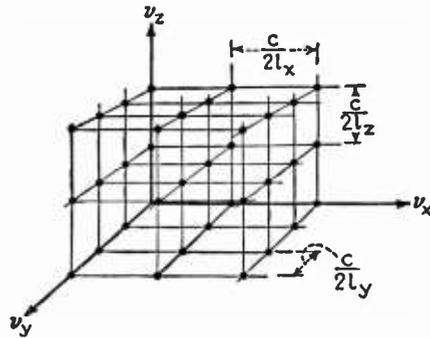


FIG. 40. Distribution of resonant frequencies for a rectangular room with side lengths l_x, l_y , and l_z . The length of a vector from the origin to each lattice point indicates the frequency and the direction of the vector indicates the direction of the corresponding standing wave. The velocity of the sound wave c is given by Eq. (1).

With the usual placement of listeners in studios and theaters the contribution of the second term is small in the audience space, and most of the intensity is due to the reverberant sound. The foregoing assumes that I is uniform throughout the room. This is reasonably valid if the room has a reverberation time near optimum, if the

Table 3. Representative Acoustical Coefficients

Manufacturer	Material	Thick- ness, in.	Absorption coefficients					
			Frequency, cps					
			128	256	512	1,024	2,048	4,096
Armstrong Cork Co.....	Cushiontone A-1	½	0.05	0.18	0.56	0.76	0.77	0.73
Armstrong Cork Co.....	Corkoustic B-5	1½	0.18	0.41	0.70	0.51	0.58	0.65
Armstrong Cork Co.....	Arrestone	1¾	0.25	0.56	0.99	0.99	0.77	0.60
Celotex Corp.....	Acousti-Celotex CS-1	¾	0.09	0.15	0.61	0.77	0.70	0.64
Celotex Corp.....	Muffitone, fis- sured	1	0.09	0.29	0.83	0.97	0.77	0.71
Celotex Corp.....	Acousteel	1¾	0.25	0.52	0.99	0.99	0.81	0.60
Johns-Manville Sales Corp.....	Fibretone	¾	0.09	0.23	0.52	0.58	0.58	0.68
Johns-Manville Sales Corp.....	Fibracoustic	1	0.25	0.62	0.72	0.72	0.71	0.76
Johns-Manville Sales Corp.....	Econacoustic MA	1¾	0.25	0.56	0.99	0.99	0.91	0.82
National Gypsum Co.....	Econacoustic	¾	0.05	0.17	0.62	0.83	0.77	0.74
National Gypsum Co.....	Acoustex 4OR	¾	0.18	0.35	0.87	0.89	0.87	0.65
National Gypsum Co.....	Acoustimetal	1¾	0.26	0.47	0.97	0.99	0.88	0.88
Owens-Corning Fiberglas Corp.	Fiberglas acous- tical tile A	¾	0.09	0.28	0.72	0.89	0.85	0.83
United States Gypsum Co....	Acoustone F	1½	0.08	0.25	0.76	0.84	0.78	0.73
United States Gypsum Co....	Auditone B	1	0.28	0.48	0.64	0.86	0.77	0.75
Description of object			128		512		2,048	
Balcony opening; soft seats.....				0.50-1.00		
Brick wall, unpainted.....			0.024		0.03		0.049	
Carpet, no pad.....			0.09		0.20		0.27	
Carpet, felt pad.....			0.11		0.37		0.27	
Draperies, 10 oz per sq yd.....			0.04		0.11		0.30	
Draperies, 18 oz per sq yd.....			0.10		0.50		0.82	
Floors, concrete.....			0.01		0.015		0.02	
Floors, linoleum.....				0.03-0.08		
Glass.....			0.035		0.027		0.02	
Plaster, smooth, on tile.....			0.013		0.025		0.04	
Wood paneling.....			0.08		0.06		0.06	
Description of object			Total absorption, equivalent area, sq cm					
Audience, seated, per person.....			1,000-2,000		3,000-4,000		3,500-6,000	
Chairs, metal or wood.....			150		170		200	
Chairs, theater, heavily upholstered.....				2,500-3,000		

sound source is not highly directional, if the room is many wavelengths long, if its proportions are neither too cubical nor too elongated, and if large flat surfaces are broken up by diffusing shapes or patches of absorbing material.

When the direct sound intensity is relatively small, Eqs. (12) and (13) may be combined to yield

$$I = \frac{P_a e^{-0.00161V/TA}}{A(1 - e^{-0.00161V/TA})} \quad \text{watts/sq cm} \quad (14)$$

This relation may be used to obtain the efficiency of a loud-speaker by measuring the intensity produced, calculating the acoustic output from Eq. (14), and comparing that with the electrical power available to the speaker from the electrical source.

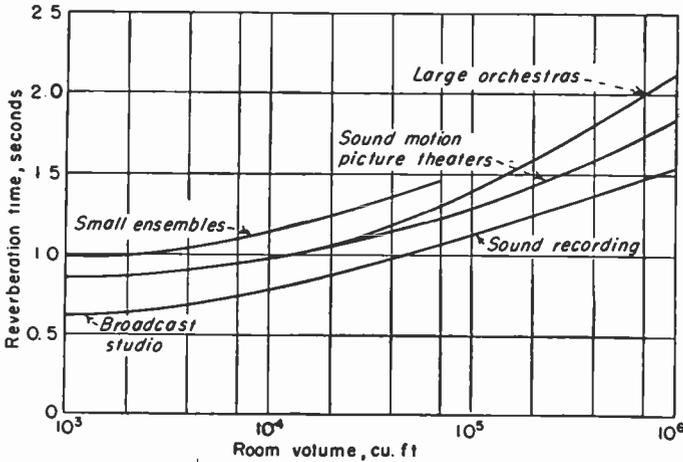


FIG. 41. Dependence of optimum reverberation time on room volume for various applications. Various authorities differ by 20 per cent or more from the indicated values so this deviation may be considered permissible.

30. Room Power Requirements. If we know the desired sound intensity, the acoustic input power P_a required to produce it may be obtained either from Eq. (13) by knowing the total room absorption or from Eq. (14) by knowing the room volume and reverberation time. If the shape of the room is unknown, the surface area A may be approximated by $SV^{2/3}$, where S is a factor between 6 for a cubic room and about 12 for an elongated room. Desirable reverberation times in terms of room volume for various uses are shown in Fig. 41. The values for speech only will lie near or below the lowest curve and are seldom realized except in acoustically treated rooms. Typical schoolrooms with average attendance, for example, usually have reverberation times well up toward the upper music range.

Speech articulation increases rapidly with intensity up to an intensity level of 40 db or 10^{-12} watt per sq cm and more slowly to 50 db or 10^{-11} watt per sq cm. If room noises are present, the speech intensity should exceed these by at least 10 db. In conversational speech the person speaking radiates about 10^{-5} watt. Loud speaking requires 10^{-3} watt. If a loud-speaker is to simulate a person speaking loudly, its acoustic output should be at least 10^{-3} watt which, for a 1 per cent efficient loud-speaker radiating all its output into the room, means an electrical input of $\frac{1}{10}$ watt.

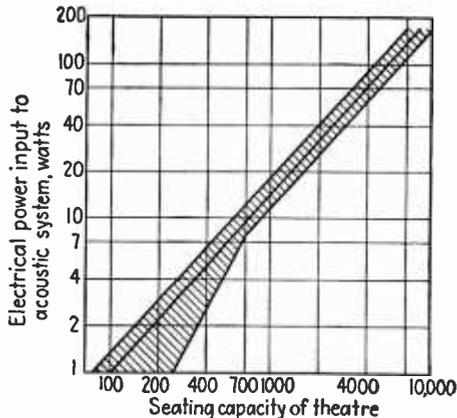


FIG. 42. Recommended amplifier output for motion-picture reproduction. Speaker system efficiency assumed to be 25 per cent, so acoustic input to theater is one-quarter indicated electrical input. Trend is toward higher values.

There is considerable difference of opinion on what constitute acceptable levels of reproduced sound. Values of electrical power input which have been suggested for theater use are shown in Fig. 42. These are based on the use of speaker systems with average system efficiencies of 25 per cent. The trend is toward larger inputs to get enhanced dramatic value in the reproduction.

In Fig. 43 are shown the *minimum* recommended acoustical power inputs to rooms having acoustical characteristics suitable for sound motion-picture theaters, for a sound pressure level of 80 db. This is satisfactory for speech in a fairly quiet room. For music reproduction, various authorities recommend inputs some 15 to 25 db higher to simulate the actual orchestra. This increases the amplifier power needed some 30 to 300 times above the recommended minimum for speech. From this it

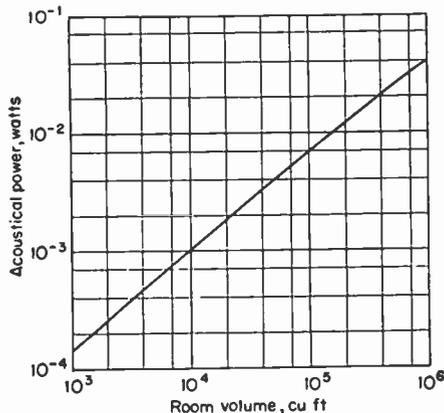


FIG. 43. Acoustical power to produce 80-db level in rooms acoustically similar to sound motion-picture theaters.

may be seen that differences of opinion may account for large performance and cost differences.

31. Acoustic System Characteristics.

When a sound receiver is included in the room with a source, we must consider the reaction of the room on it. The most common receiver is a listener. Because of the difficulty, however, of making objective measurements of what is going on in the listener's central nervous system, it is more convenient, although only approximately correct, to substitute one or more microphones for the listener.

For sake of simplicity, assume we have a loud-speaker as a source and a microphone as a receiver or sink, each with two accessible terminals. Since no source of energy is assumed in the room, these four leads may be considered the terminals of a passive quadripole or four-

terminal network. From circuit theory we know that the measurement of three independent quantities will completely specify the performance of this quadripole at any one frequency. By analogy with the purely electrical case we may define the ratio (complex) of the current through a specified microphone load to the input voltage of the speaker as a transfer admittance.

A curve giving the magnitude of this quantity for a speaker and a microphone mounted in diagonally opposite corners of a rectangular 18- by 20- by 11-ft room is shown in Fig. 44. This transfer admittance is what might be called the response curve of the loud-speaker measured in this room with designated locations for the source and microphone and with the particular microphone employed. As would be expected from the complicated equivalent circuit of the room, this transfer admittance varies by a large factor with frequency and exhibits a large number of maximums and minimums at high frequencies. In passing it may be noted that the impedance of the boundary of this room varied appreciably with frequency and therefore the location of the resonant and antiresonant frequencies of the "electroacoustoelectrical network" do not occur at the frequencies predicted by constant boundary impedances.

Considering the matter in this light, we see that this response curve depends on the type of speaker, microphone, their location in the room, the geometry of the room, and the impedance of the room boundaries and therefore the impedance of the entire dynamical system seen from the acoustic terminals of the speaker and microphone.

The loud-speaker supplies maximum energy to the room when the impedance seen by the diaphragm looking into the room is the conjugate of the impedance looking

back into the diaphragm with the speaker connected to its generator or amplifier. Since the latter impedance is usually high, the speaker supplies maximum power when the room impedance is high, which occurs (by definition) when the ratio of the pressure to particle velocity is high, *i.e.*, when the speaker is near a pressure maximum. A pressure-actuated microphone gives maximum response at a pressure maximum. The maximum peaks in the transfer admittance of Fig. 44 therefore occur when both the microphone and speaker are near pressure maxima. No resonance pressure maximum occurs at the speaker below the lowest resonant frequency of the room, and good l-f response is therefore hard to obtain in small rooms.

The apparent damping or Q of a mode of the system may be obtained by steady-state sharpness of resonance methods corresponding to those employed in circuit investigations. With negligible dissipation due to losses in the source, air, and sink, the damping of a particular mode is an indication of the absorption of energy at the boundary. From this damping coefficient the effective absorption coefficient of the boundary under the conditions determined by the mode may also be obtained. The

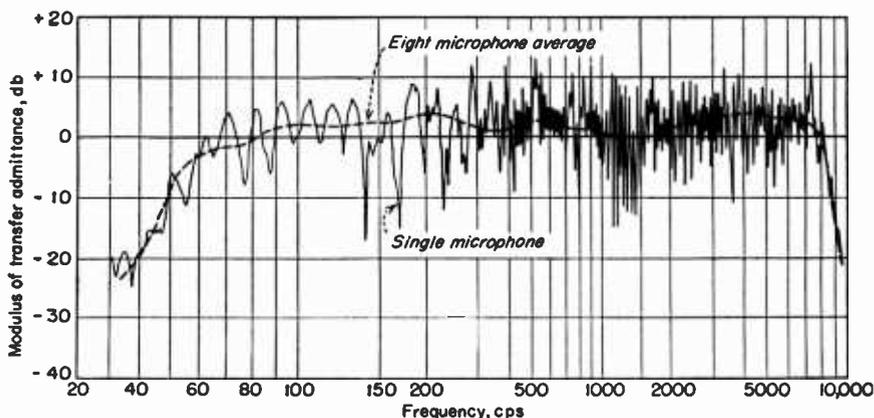


FIG. 44. Transfer admittance or "response curve" of a speaker, room, and microphone or electroacoustoelectrical transducer. Speaker and microphone are in diagonally opposite corners of an 18- by 20- by 11-ft room.

simplest result is obtained when the wave front is parallel to the walls on which it is incident. The same apparent absorption is obtained from the transient decay of the same mode.

Likewise, by analogy with the electrical case, we may think of the transient current which flows through the microphone load when a unit d-c potential is suddenly supplied to the speaker terminals as the transfer indicial admittance of this electroacoustoelectrical network. Viewed in this light the transient response of the speaker itself (a small part of the dynamical system) or the transfer indicial admittance of the loud-speaker and microphone mounted in a free field where no reflections are present takes on much less significance.

We know experimentally that any room which is considered a good acoustical listening environment has appreciable reverberation and therefore that the rate of decay of the energy in the resonators of the room is only moderate. Experimentally it has been found that the rate of decay of the modes of the speaker itself measured in a free field is of the same order. If the loud-speaker is loosely coupled to the room, *i.e.*, if the room impedance seen by its diaphragm is small as compared with the impedance seen looking into the diaphragm, then we may loosely think of the loud-speaker as converting the unit d-c emf into a number of damped sinusoidal terms (one corresponding to each mode of the speaker), which in turn excite the room. The

spectral composition of a single damped sinusoid for two rates of decay is shown in Fig. 7. From these we see that, if the rate of decay is large, the number of room modes excited may be large, because of the broad frequency spread of the energy exciting the room. Actually, of course, we should only think of the dynamical system as a whole and the above explanation as a simplification of the problem.

Experimental curves of this transfer indicial admittance are what one would predict from the theory. If the speaker is loosely coupled to the room, if its damping when it looks into a small acoustical impedance is low, and if an undamped resonant frequency of the speaker lies near one of the resonant modes of the room, the transient term looks like a typical one for two coupled circuits. That is, there are two prominent rates of decay containing the combined effect of the two important resonant frequencies (the speaker and room). On the other hand, if the speaker is highly damped when looking into a high acoustical impedance, if the driving point impedance of the room at an undamped natural frequency of the diaphragm is high, and if the room is large or its dimensions are so chosen that there are a number of resonant frequencies near an undamped resonant frequency of the speaker diaphragm, then the transient term consists of the superposition of a large number of damped sinusoids. In this case it may be seen that elimination of the term due to the loud-speaker would cause a negligibly small change in the apparent transient response of the system. This was verified in an unreported investigation in which it was shown experimentally that, if the fundamental speaker mode was eliminated by the use of a properly chosen electrical network, the aural result in reproduced speech and music was small unless the damping of the speaker radiating into a free field was unusually small.

The more important practical implications of the above (see also Sec. 7, Mutual Radiation Impedance) are the following: (1) The loud-speaker should preferably be mounted in the corner of the room. In this position the greatest number of room resonances are "excited" and the most energy is supplied to the room. (2) The *average* l-f radiation is a maximum when the speaker is as near the floor (or ceiling) as possible and in the room corner. Next most desirable location is near floor (or ceiling) and side wall. (3) At any one frequency, maximum radiation is obtained when the room impedance seen by the diaphragm is the conjugate of the impedance seen looking back into the diaphragm. That is for the location which makes the *combined* speaker and room resonate. Such a maximum may not be obtained if the longest room dimension is less than roughly a half wave length long. (4) Because of this and the small number of resonant frequencies which occur in small rooms at low frequencies, small rooms do not normally permit the best l-f reproduction. (5) Corner positions also permit improved h-f response because of the smaller solid angle the radiation has to cover. (6) The l-f transient response of the speaker itself is not so important as is generally supposed because the transient response of the room helps obscure this distortion.

OBJECTIVE LOUD-SPEAKER TESTS

The following more important characteristics of a loud-speaker must be determined in any complete test: response-frequency, efficiency-frequency, directional, impedance, and distortion.

32. Response-frequency Characteristic (Steady State). A response-frequency curve of a speaker is a curve graphically depicting the sound produced at a designated position in the medium, the electrical input and acoustic environment being specified. Frequency discrimination is the most important form of distortion in many loud-speakers, and the response curve attempts to indicate quantitatively the amount present. Since the ear is primarily responsive to the sound pressure, the ordinate of the curve is made proportional to it or to its average value in a specified region.

The response curve is obtained by connecting the loud-speaker to a variable frequency source of specified internal impedance and constant specified internal voltage. The pressure at one or more points in the medium is measured as the frequency is

varied slowly enough so the resulting measurement does not differ appreciably from the steady-state value.

A "free-field" response curve is made outdoors in the absence of unintended reflecting surfaces and is probably the most useful single curve showing the loud-speaker performance for outdoor applications. Curves of this type are valuable because (1) the direct incident sound from the source in various directions may be accurately determined and a close estimate made of the direct sound indoors; (2) the acoustic environment is relatively simple since only intended reflecting surfaces are included. The efficiency of the unit may then be accurately obtained for this environment at some frequencies and estimated at others, since the impedance seen by the diaphragm will change slowly with frequency, *i.e.*, the transfer admittance of the speaker, air, and microphone is a smooth curve. (3) The specified test conditions may be duplicated relatively easily at various laboratories permitting significant comparison of test results. The construction of identical test rooms, however desirable, would be difficult, partly because of differences of opinion on an "average" room and partly for economic reasons.

By "intended" reflecting surfaces is meant those that are an intended part of the radiating system. Frequently a cabinet or enclosure is measured outdoors in the absence of all reflecting surfaces, *i.e.*, radiating into a solid angle of 4π steradians or a complete sphere. This is usually undesirable since most enclosures are intended to operate on a floor and against a wall which adds two intended reflecting surfaces and primary images (see Radiation Impedance). If the enclosure is intended to operate in a room corner, there are three important primary images. The impedance seen looking into these reduced solid angles of π and $\pi/2$ steradians, respectively, is very different from the 4π case and usually results in an error of the order of 6 to 8 db at low frequencies. The error is especially large in the case of compound sources such as vented enclosures. The measured nonlinear distortion usually differs by a much larger factor.

Outdoor measurements into solid angles of π and $\pi/2$ steradians are made by constructing large rigid nonabsorbing surfaces.

Other intended parts of the speaker such as the baffle, horn, and enclosure should, of course, be specified. The normal impedance or the impedance looking into the signal terminals of the speaker with the acoustical load (acoustical environment), used when the response curve was obtained, should be plotted. Both the angle and modulus of this impedance are required if the response of the speaker with any source impedance other than that employed in the test is to be calculated.

If a space average of the pressure is obtained by moving the microphone or by using multiple microphones, details of the method should be given. If a warble tone or noise generator is used to get a "moving frequency average" of the transfer admittance, the spectral composition of the source should be specified. These expedients and the one involving motion of the loud-speaker, which is usually unsatisfactory, are recommended only for indoor measurements when the room does not provide approximately free-field conditions.

Normal listening-room measurements are made with the loud-speaker mounted in its intended position in a typical listening room. As noted under Room Acoustics, the room impedance seen by the loud-speaker depends on the characteristics and location of the source itself (diaphragm sizes, locations, and modes of vibration), the geometry of the room, and the impedance of its boundaries. This means that the energy supplied the room depends on the particular room and speaker location chosen. This is frequently used as an argument against this type of test. Since the results obtained in reasonably similar rooms, with similar speaker locations in each, differ by only a moderate amount, this disadvantage does not outweigh the many important advantages of this type of test, some of which are (1) the impedance seen by the loud-speaker (including cabinet or enclosure), averaged over a small frequency interval, is closer to the average impedance seen under operating conditions than the impedance seen under the usual (4π steradians) outdoor test conditions; (2) calculation of the

indoor from the outdoor performance is only of academic interest when hundreds of response curves are to be obtained, because of the labor involved; (3) ready comparison of the results of objective and subjective or listening tests in the same room is possible if the room is a good listening room; (4) one is not at the mercy of the weather; and most important (5) test facilities are readily provided in almost any organization.

Three large laboratories measured one speaker and plotted what they would publish as the response-frequency graph of the speaker. The results are shown in Fig. 45.

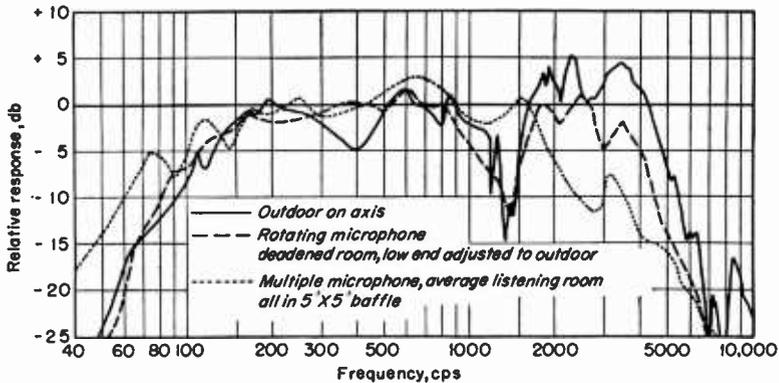


FIG. 45. Response of one speaker as measured by three different companies, showing the futility of relying on response curves without a complete knowledge of the specific test, acoustic environment, and method.

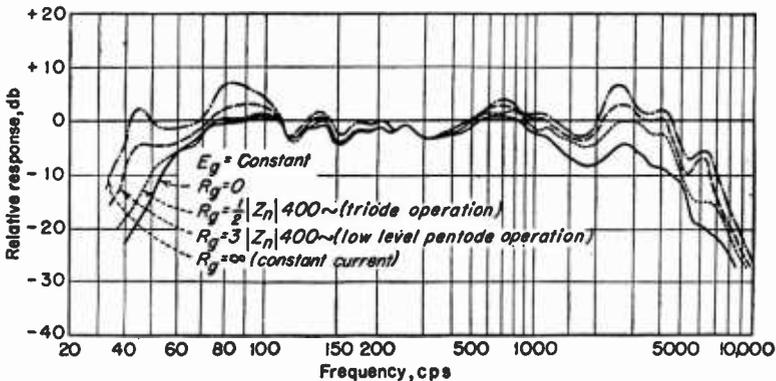


FIG. 46. Variation in total sound power radiated by a bass reflex enclosure into a room as source impedance (impedance match) is varied. Generator or source voltage constant for each curve but arbitrarily raised as the source resistance was raised to maintain constant acoustic output at 400 cycles.

This does not indicate any error in measurement. Actually different things were measured in each case. The curves indicate that response curves must be interpreted with great care and then only by a person familiar with the many factors involved. No experienced speaker designer or user thinks of choosing a speaker solely or even largely on the basis of a response curve.

33. Efficiency-frequency Characteristic. If the free-field-pressure response at a sufficient number of points on a spherical surface centered on the diaphragm is obtained, the total acoustical output may be calculated.

The efficiency-frequency curve of a speaker corresponds to the response-frequency curve except that the ordinate indicates the efficiency (usually "absolute" or system). In a typical listening environment and listener location the direct incident sound energy, which would be approximately indicated by the free-field response-frequency curve at the listener's location with respect to the speaker, is only a small fraction of the reflected sound energy. A curve which gives the pressure, averaged over the useful listening region, then indicates the probable pressure the listener will experience. If absorption at the room boundary is independent of frequency this will be proportional to the total energy emitted by the speaker. Efficiency-frequency or space-averaged response-frequency curves are therefore the most useful in interpreting indoor operation. Outdoor response-frequency curves at various angles off the speaker axis, with the speaker radiating into approximately the solid angle it will see indoors, are also desirable since the listener, owing to his ability to localize sounds, weights the direct incident sound energy particularly at high frequencies more heavily than the same energy if in a reflected wave.

SUBJECTIVE LOUD-SPEAKER TESTS

34. Listening or Subjective Tests. Listening tests are a necessary part of the complete test of a loud-speaker. While objective measurements are valuable in design work and in quantitatively determining some performance criteria, they cannot at present completely specify the subjective performance. Because of the apparent simplicity of listening tests many important factors are frequently neglected, with the result that many tests are meaningless and others actually misleading.

In both indoor and outdoor tests all precautions should be taken that are used in objective tests. The only essential difference is that the listener is substituted for the microphone. The properties of the ear and listener must therefore be considered in interpreting the results.

35. Relative-loudness Efficiency. The most common test is one to determine the relative-loudness efficiency of two speakers. An attenuator in the amplifier which does not alter its response is adjusted (usually with a relay which also switches the speakers) to attenuate the input to the louder speaker by the amount required to make the speakers equally loud. The required attenuation of the louder in decibels is their relative loudness efficiency in decibels. The relative loudness will depend primarily on the spectral composition of the test signal, the response-frequency characteristic of the speakers, and on the sound intensity. Tests on the speaking and singing voice and various types of music are usually averaged. A valuable signal source for this and response-frequency tests is a "flat" noise source, or one in which the energy is uniformly distributed with frequency. This particular spectral composition ensures energy at frequencies at which significant differences in the speaker response may occur.

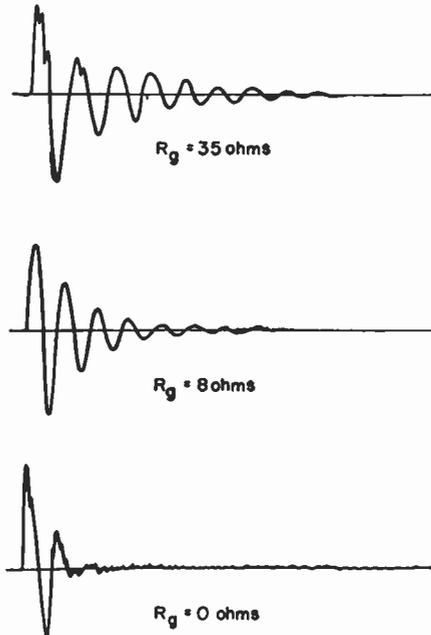


FIG. 47. Response of moving-coil speaker with 8-ohm voice coil to an electrical pulse as source resistance is varied. Minor irregularities in curves are due to h-f modes of vibration of diaphragm.

36. Response-frequency Characteristic. Apparent subjective response-frequency tests may be made with the same signal sources used in the loudness tests. Since the listener is not mobile, "space-averaging" methods employed with microphones cannot be used and "frequency-averaging" methods are employed. While noise sources are occasionally used in objective tests, they have unfortunately been neglected in subjective tests, where they are of special value because the trained ear can quickly appraise response differences which are missed if the signal source contains no energy at the frequencies at which differences occur.

37. Distortion Characteristic. Except with a single- or double-frequency input (the latter to determine intermodulation) it is difficult to determine the distortion characteristic of the speaker itself. With one or two simultaneously applied frequencies the input to the speaker is readily determined when the normal impedance of the speaker is known. This is not true of a signal of random energy distribution, and therefore with such a signal the apparent input to the speaker is not readily determined unless the normal impedance is relatively independent of frequency. When the *speaker* distortion characteristic is desired, the amplifier should be capable of supplying many times the rated input power to the speaker without distortion because of the high ratio of peak to average energy in speech and music (see these sections). Much overload charged to speakers is amplifier overload.

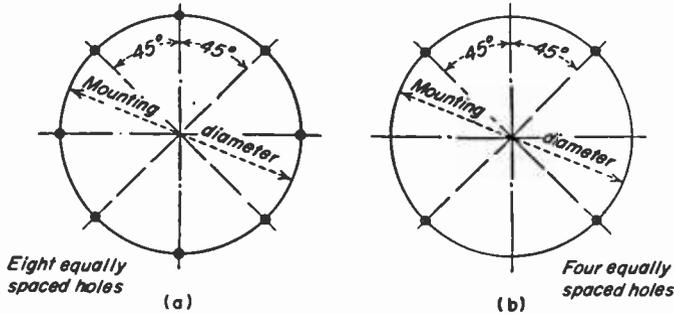


Table 4. RMA Standard Mounting Dimensions for Direct-radiator Speakers

Nominal speaker size, in.	Hole arrangement	Mounting diam, in.	Min hole diam, in.	Baffle* hole diam, in.
3½	b	3½ ¹ / ₈	¾ ¹ / ₈	3½
4	b	4½ ¹ / ₈	¾	3¾
5	a	4½ ¹ / ₈	¾	4½
5¾	a	5¾	¾	4¾
6½	a	6½	¾	5¾
8	a	7¾	7⁄8 ² / ₈	6¾
10	b	9¾	7⁄8 ² / ₈	8¾
12	b	11¾ ¹ / ₈	¾	10¾
15	a	14¾ ¹ / ₈	1¾ ¹ / ₄	13¾

* Not an adopted standard.

What is usually measured is the *system* distortion characteristic. Since amplifier overload almost invariably occurs at about the level at which speaker overload occurs in an economically planned system, what is measured is the combined system overload. In this case the speaker with the most restricted h-f response (other factors being equal) will have the best *system* overload rating since annoying h-f distortion products are attenuated.

There is no standard for speaker input power rating, but in practice a speaker rated at X watts will "handle" the output of an X-watt amplifier, which is not overloaded,

with speech or music (complex wave) input (see Motors and Permanent and Electromagnets). The rating does not indicate the power the speaker will handle at a single frequency but takes advantage of the *normal* (no heavy bass or treble accentuation) spectral composition of speech and music. To avoid the trouble in determining the signal level across the speaker, with its variation in normal impedance, the grid voltage on the output stage may be measured. The signal input is raised until perceptible distortion results. The peak grid voltage is then measured with an indicator having a negligible time constant, such as a cathode-ray tube. A resistance equal in value to the magnitude of the nominal loud-speaker impedance is substituted for the speaker. The power dissipated in the resistance load with a 400-cycle signal having the same maximum value as the maximum signal is the system (since it includes the output stage) speaker input power rating.

38. Outdoor Tests. These should be conducted so the listener subtends the same or a known angle with each source. The sources should be mounted so their mutual-radiation impedance does not influence the result. Separating the sources by several times the diameter of the cone or horn mouth usually suffices. The energy absorbed by the unused speaker will be more nearly independent of frequency if the voice coil of the unused speaker is short-circuited. Unwanted reflecting surfaces should be avoided.

39. Indoor Tests. The speakers should be separated by several times the diameter of the cone or horn mouth to minimize mutual-radiation impedance. This is particularly true if the speakers are mounted on a common open baffle. Some coupling between the sources will always exist because of the transfer admittance between the two in the room (see Room Acoustics). It is important to mount the speakers symmetrically with respect to the room and listener in order to provide similar coupling between each source, the room, and the listener.

References

General:

- American Standards Association, Z24.1-1942, Z24.2-1942, Z24.3-1944, Z24.4-1942, C16.4-1942.
Bell System Tech. J. (see Cum. Index, 1-10).
 Crandall, I. B.: "Theory of Vibrating Systems and Sound."
Electronics (McGraw-Hill Publishing Company, Inc.)
 Geiger, H., and Karl Scheel: "Handbuch der Physik," vol. 8, Akustik, 1927.
 IRE Electroacoustic Standards, 1938 (also *Proc. IRE*).
J. Acoust. Soc. Am. (see Cum. Index, 1-10).
 McLachlan, N. W.: "Loudspeakers."
 Morse, P. M.: "Vibration and Sound."
 Olson, H. F.: "Elements of Acoustical Engineering," 2d ed., 1947.
 ——— and F. Massa: "Applied Acoustics," 1939 ed.
 Rayleigh, Lord J. W. S.: "Theory of Sound."
Soc. Mot. Picture Engrs., J.
 Stewart, G. W., and R. B. Lindsay: "Acoustics."
 Wien-Harms: "Handbuch der Experiment Physik," Vol. 17, Parts 1-3.

Speech, Music, and Hearing:

- Beasley, W. C.: Characteristics and Distribution of Impaired Hearing in the Population of the United States, *J. Acous. Soc. Am.*, **12**, 114-121, 1940; also National Health Survey, *Hearing Study Series Bulls.* 1-7.
 Fletcher, H.: "Speech and Hearing"; also Physical Characteristics of Speech and Music, *Bell System Tech. J.*, **10**, 349, 1931; also Hearing, the Determining Factor in High Fidelity Transmission, *Proc. IRE*, **30**, 266, 1940; Steinberg, J. C.: Articulation Testing Methods, *Bell System Tech. J.*, **10**, 806, 1929; and Munsen, W. A.: *J. Acous. Soc. Am.*, **5**, 83, 1933.
 Sivian, L. J.: Speech Power and Its Measurement, *Bell System Tech. J.*, **10**, 646, 1939; and Dunn, H. K., and S. D. White: "Absolute Amplitudes and Spectra of Musical Instruments and Orchestras."
 Snow, W. B.: Audible Frequency Ranges of Music, Speech and Noise, *J. Acous. Soc. Am.*, **3**, 155-166, 1931.
 Stevens, S. S., and H. Davis: "Hearing, Its Psychology and Physiology."

Loud-speakers:

- Bostwick, L. G.: *Bell System Tech. J.*, **8**, 135, 1929. (Tests.)
 Cook, E. D.: *Gen. Elec. Rev.*, **33**, 505, 1930. (Tests.)
 Greaves, V. F., F. W. Kranz, and W. D. Krozier: The Kyle Condenser Loudspeaker, *Proc. IRE*, **18**, 1142, 1929.

- Hall, W. M.: *J. Acous. Soc. Am.*, **3**, 552, 1932. (Horns.)
- Klapman, S. J.: Interaction Impedance of a System of Circular Pistons, *J. Acous. Soc. Am.*, **11**, 289, 1940.
- Knowles, H. S.: *Electronics*, **4**, 154, 1932; also **6**, 240, 1933. (Tests.)
- Olney, B.: *Proc. IRE*, **10**, 1113, 1931. (Tests.)
- Pedersen, P. O.: *J. Acous. Soc. Am.*, **6**, 227, 1935; and **7**, 64, 1935. (Subharmonic theory.)
- Phelps, W. D.: *J. Acous. Soc. Am.*, **12**, 68, 1940. (Horn losses.)
- Rice, C. W., and E. W. Kellogg: *J. AIEE*, **44**, 982, 1015, 1925. (Moving-coil speaker.)
- Salmon, V.: *J. Acous. Soc. Am.*, **17**, 199, 212, 1946. (Horn theory and new horns.)
- Stensel, H.: *Elektrische Nach. Tech.*, **4**, 239, 1927; **6**, 165, 1929; and **7**, 90, 1930; also *Z. Tech. Physik*, **10**, 569, 1929; and *Ann. Physik*, **11**, 947, 1930. (Directivity of radiators.)
- Wegel, R. L.: *J. AIEE*, **49**, 791, 1921. (Theory.)
- Wolf, I., and L. Malter: *Phys. Rev.*, **33**, 1061, 1929. (Mutual radiation impedance.) Also *J. Acous. Soc. Am.*, **2**, 201, 1930. (Directivity.)

Room Acoustics:

- Acoustical Materials Association Bulletins.
- J. Acous. Soc. Am.*, Symposium, July, 1940, Bibliography (see also Cum. Index, 1-10).
- Knudsen, V. O.: "Architectural Acoustics," 1932.
- Morse, P. M., and R. H. Bolt: Sound Waves in Rooms, *Revs. Modern Phys.*, **16**, 69, 1944.
- Sabine, P. E.: "Acoustics and Architecture," 1932.
- Watson, F. R.: "Acoustics of Buildings," 1930; and *J. Acous. Soc. Am.* **3**, 14-43, 1931.

CHAPTER 17

RECEIVING SYSTEMS

BY CHARLES E. DEAN¹

1. **Classification.** Radio receivers may be classified² according to their operating principles:

1. Tuned radio frequency.
2. Regenerative.
3. Superheterodyne.
4. Superregenerative.

Receivers of all these types have been made and used for amplitude-modulated (a-m) signals. For frequency modulation (f.m.) the superheterodyne is almost always used.

The following sections give general descriptions of these types, after which the various types and related topics will be discussed in detail.

In the discussion of receivers in the present chapter, reference is often made to the various frequency bands by the following abbreviations³:

2. Tuned-radio-frequency Receivers

Tuned-radio-frequency (t-r-f) receivers obtain their selectivity and r-f amplification solely through the use of selective circuits and tubes which function at the frequency of the incoming signal. This type has been largely replaced by the super-

Abbreviation	Meaning	Frequency range
v-l-f	Very-low-frequency	10-30 kc
l-f	Low-frequency	30-300 kc
m-f	Medium-frequency	300-3,000 kc
h-f	High-frequency	3-30 Mc
v-h-f	Very-high-frequency	30-300 Mc
u-h-f	Ultra-high-frequency	300-3,000 Mc
s-h-f	Super-high-frequency	3,000-30,000 Mc
e-h-f	Extremely high-frequency	30,000-300,000 Mc

heterodyne. The t-r-f receiver uses two or more tuned circuits, which are usually adjusted simultaneously by a single gang capacitor tuning control.

The series resistance of a conventional tuned circuit, whose frequency is varied by means of a variable capacitor, increases with frequency. This corresponds to a constant value of Q . The band width of t-r-f receivers therefore varies in proportion to frequency, being about three times as great at 1,600 as at 550 kc. One or two of the tuned circuits in a t-r-f receiver are generally used in the antenna-input system, and the remainder provide the coupling between the r-f stages. One or two stages of a-f amplification are used in the audio portion of the receiver.

3. **Regenerative Receivers.** In a regenerative receiver, the received voltage is impressed on the grid of a vacuum tube; a portion of the resultant amplified voltage is fed back to the grid in the proper phase relation to increase the applied grid voltage.

¹ Hazeltine Electronics Corporation, Little Neck, L. I., N. Y.

² For an elementary treatment see, Jordan, Nelson, Osterbrock, Pumphrey, and Smeby, "Fundamentals of Radio," edited by W. L. Everitt, Prentice-Hall, 1942; Keith Henney, "Principles of Radio," 5th ed., Wiley, 1945; David Grimes, "Meet the Electron," Pitman, 1944; M. G. Sufferin, "Basic Radio Principles," McGraw-Hill, 1943. Certain general reference books containing material on various aspects of radio receivers are listed at the end of this chapter.

³ Report of Standards Committee Relating to Standard Frequency Ranges, *Proc. IRE*, August, 1945, p. 548. For another system, see Proposed Frequency Band Designations, *Elec. Eng.*, May, 1947, p. 471.

This action reduces the effective resistance of the resonant circuit where the signal is applied and thereby provides considerable amplification of the received signal.

Regenerative receivers are usually provided with controls for tuning and for controlling the feedback. If the feedback is increased beyond a certain value, sustained oscillations will result and the beat frequency produced between the carrier of the transmitter and the locally produced oscillations indicates when the receiver is properly tuned. This method of tuning is called "zero beat," since, as the tuning of the receiver is adjusted, the beat note decreases in frequency until it is no longer audible. When a conventional regenerative receiver is tuned in this way, interference is produced in nearby receivers tuned to the same station. A stage of amplification between the antenna and the regenerative circuit will reduce the likelihood of this interference. The regenerative receiver is quite sensitive considering the small number of tubes employed; as only a single tuned circuit is generally used, the selectivity is limited. Regenerative receivers are still used to a limited extent in amateur and other h-f services.

4. Superheterodyne Receivers. In the superheterodyne the signal voltage is combined with a local oscillator voltage and converted into a new signal with the same modulation but a lower ("intermediate") carrier frequency, which is then amplified and detected to reproduce the original modulation.

The superheterodyne has the essential components of a t-r-f receiver, but in addition it has a frequency converter and an i-f amplifier. The frequency converter consists of an oscillator and a modulator. The oscillator and the t-r-f circuits are usually tuned simultaneously by means of a gang capacitor or a gang inductor. Constant frequency difference is maintained between the oscillator and r-f circuits, through the use of either (1) a combination of fixed shunt and series capacitors in the oscillator circuit in conjunction with a gang capacitor in which all the sections are identical, or (2) a gang capacitor in which specially shaped plates are used in the oscillator section, or (3) suitable arrangements of variable-inductance tuning.

The i-f amplifier in broadcast receivers generally uses two or three transformers, each containing two coupled tuned circuits. With this substantial number of circuits and the fixed and favorably chosen i-f value, the i-f amplifier provides the major portion of the amplification and selectivity of the receiver. Since the characteristics of this amplifier are independent of the frequency to which the receiver is tuned, the sensitivity and selectivity of a superheterodyne are usually fairly uniform throughout its tuning range. The r-f circuits are used primarily for eliminating certain forms of interference common to this receiver. The performance of the superheterodyne is in general superior to that of any other type of receiver in use today. The most common i-f value in a-m broadcast receivers is 455 or 456 kc.

5. Superregenerative Receivers. A superregenerative receiver is a regenerative receiver in which sustained oscillations are prevented by the periodic variation of the effective resistance of the resonant circuit to which the received signal is applied.

In the superregenerator, oscillations are permitted to build up repeatedly in a resonant circuit tuned to the received signal. Sustained oscillations are prevented by the application of a quenching-frequency potential to the grid of the superregenerative tube. This potential periodically affects the tube characteristics in such a way as to stop the oscillations. The quenching frequency may be supplied either by a separate oscillator or by the superregenerative tube itself. If this frequency is in or near the upper part of the a-f range, the audio system of this type of receiver is usually provided with an a-f filter to remove the quench frequency from the audio output. An r-f stage ahead of the detector will prevent energy from being transferred from the superregenerative circuit to the antenna. A signal input of 50 to 100 μv will give an intelligible signal although an input of 500 to 1,000 μv is generally necessary to reduce the noise to a satisfactory value. Harmonics of the quench frequency beating with the received signal make interference if the ratio between signal and quench frequencies is not 100:1 or more.

The superregenerator was used extensively to receive pulse-type signals in the v-h-f

band during the Second World War and is likely to come into wider use now that its advantages are becoming better understood.

6. **F-m Receivers.** During recent years, f.m. has come into extensive use in broadcast, police, military, and other services. The carrier frequency can be varied, if desired, over a wider range than the extent of the signal components. For example, in broadcast service in the United States a set of standards has been established by the FCC, under which audio frequencies to 15 kc are handled, but the maximum swing of frequency is ± 75 kc.

The usual broadcast f-m receiver is a superheterodyne differing in the following respects from m-f and h-f a-m broadcast receivers: (1) the carrier frequencies are much higher; (2) the band width is wider to accommodate the wide swing of the signal and the drift of the local oscillator; (3) the i.f. is higher, the present standard value being 10.7 Mc; (4) the i-f amplifier may contain one or more limiter stages; (5) the a-m detector is replaced by a frequency-sensitive device to reproduce the original audio wave; (6) the audio circuit contains a deemphasis circuit to remove the preemphasis inserted at the transmitter, this feature being another antinoise measure; and (7) the audio amplifier and speaker should be (but often are not) high-fidelity types to realize the high quality of which the system is capable.

7. **Method of Rating.** Receivers are generally rated on the basis of the following characteristics:

1. **Sensitivity** is that characteristic which determines the minimum signal strength to which the receiver will respond. It is measured quantitatively in terms of the input voltage or power required to give a standard output. Therefore, the receiver with the smallest value of this input is the most sensitive.

In the m-f and h-f bands the necessary input is usually stated in microvolts or in the corresponding number of decibels below 1 volt. For example, a set might have a sensitivity of $10 \mu\text{v}$ or 100 db.

The statement of sensitivity in terms of power is on the basis of *available power*, which is $E^2/4R$, where E is the open-circuit voltage of the signal generator in the sensitivity measurement and R is the resistance of the dummy antenna and the signal generator which together act as the source of signal supplying the receiver. The expression of sensitivity in terms of power is the usual practice in the v-h-f and higher bands.

2. **Noise** produced by a receiver, after defects such as hum, poor contacts, and faulty parts are corrected, comes from thermal noise in the antenna resistance and from noise originating in the tubes. Such noise is not large enough to be important in broadcast-band receivers having medium sensitivity and provided with an r-f stage, because high voltage gain in the first coupling network and first tube increases the signal to a value great enough to make all sources of noise small in comparison. Noise may be an important factor in the case of sets not having an r-f stage, and also in the case of all receivers requiring high sensitivity, such as automobile and farm sets.

In the v-h-f and higher ranges of frequency the input resistance of tubes is so low that the early circuits of the receiver cannot raise the signal voltage high enough to make the noise negligible. The amount of resistance noise and tube noise originating within the set is therefore important, and a quantity called the "noise figure" is used as a measure of this quantity. This is described in Sec. 20.

3. **Selectivity** is the degree to which the receiver is capable of differentiating between the desired signal and signals of other carrier frequencies. This characteristic normally requires one or more graphs for its expression.

4. **Fidelity** is the degree to which the receiver accurately reproduces at its output terminals the modulation possessed by the received wave.

5. **Maximum undistorted output** for a broadcast receiver is defined as the maximum power output which can be obtained from it when the output voltage does not contain more than 10 per cent of total harmonics.

METHODS OF TESTING

Standard methods of testing radio receivers for the reception of a-m signals in the m-f and h-f bands have been established by the IRE and are described in detail in "Standards on Radio Receivers: 1938." Very similar treatment of the more important tests, as made by receiver producers, is given in "The Measurement of Perform-

ance Characteristics of Amplitude Modulated Radio Receivers," of the RMA Data Bureau, December, 1946. The following paragraphs give some of the essentials of the methods established by IRE and RMA.

8. Definition of Terms.

1. *Sensitivity, selectivity, fidelity, and maximum undistorted output* were defined in Sec. 7.
2. *Normal Test Output.* An a-f power output of 0.5 watt in a standard dummy load connected across the output terminals of the receiver is the normal test output of a home receiver except when the maximum power output is less than 1 watt and more than 0.1 watt, in which case the normal test output is 0.05 watt. The 0.5-watt value is assumed unless the 0.05 value is stated. For automobile broadcast receivers, the normal test output is 1.0 watt.

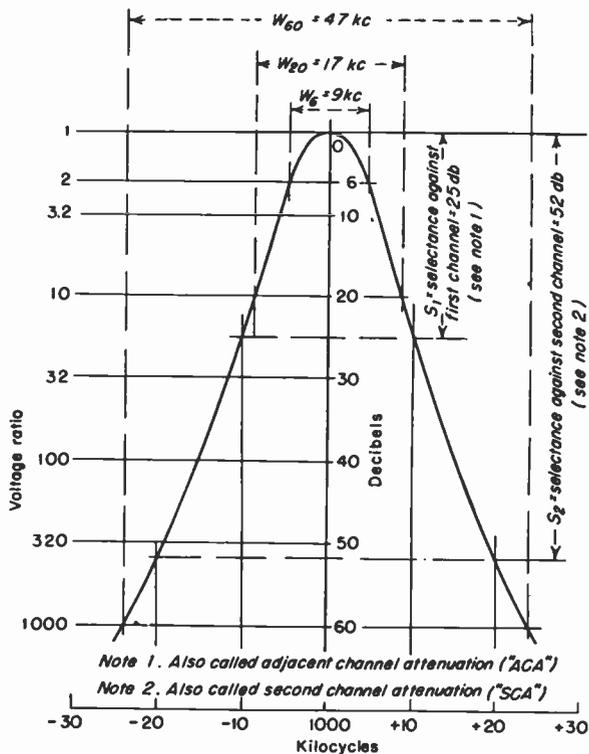


FIG. 1. Definition of selectance and band-width values on typical selectivity curve of a complete receiver.

3. *Sensitivity-test Input.* This is the rms r-f voltage, modulated 30 per cent at 400 cycles, which results in normal test output at resonance. It is applied through a standard dummy antenna.

4. *Selectance* is the selectivity against interference on a definite channel near the one to which the receiver is tuned. As shown in Fig. 1, it is expressed in decibels, or it may be given as the corresponding voltage ratio. The selectance against the next channel (10 kc from resonance) is also called the adjacent-channel attenuation (ACA), and against the next channel beyond (20 kc from resonance) is called the second-channel attenuation (SCA). The values should be averages of the decibel figures for the particular frequencies above and below the resonant frequency, or the corresponding voltage ratio obtained by taking the geometric mean of the two observed voltage ratios. Notation to designate each observation has been established, S_{+1} representing the selectance for the first channel above

the tuned frequency and S_{-1} the first channel below. The letter S_1 represents the mean value computed as described. Likewise S_{+2} , S_{-2} , and S_2 represent values of the SCA.

5. *Band width* is another means of expressing selectivity and is used extensively. It is the total width of the frequency band at the edges of which the attenuation has a specified value. In many laboratories the notation W_6 , W_{20} , W_{40} , and W_{60} , is used to represent the band width at 6, 20, 40, and 60 db, respectively.

9. Receiver Test Equipment.

1. As a source of voltages for testing, a signal generator is necessary. This is a shielded oscillator whose frequency, for testing a-m receivers, can be varied from 100 to 30,000 kc. An a-f oscillator is provided to modulate the r-f oscillator by a known amount at any frequency from 30 to 10,000 cycles. For most tests 400-cycle modulation is used. A calibrated attenuator is provided to furnish a range of outputs from 0.3 μ v to at least 100,000 μ v and preferably to 2,000,000 μ v (from 130 db below to a preferable value of 6 db above 1 volt).

For the v-h-f and higher ranges, special signal generators¹ are required. A thermistor output meter has been used with a mutual-inductance wave-guide attenuator.

2. *Standard Dummy Antenna.* The standard dummy antenna for a broadcast radio receiver is shown in Fig. 2, which is taken from the 1938 IRE receiver standards. It simulates typical actual outdoor antennas from 150 to 30,000 kc. The resistive and reactive components of the impedance are shown, as well as the modulus, which is the square root of the sum of the squares of the two components. At the higher frequencies the reactance can be seen to become negligible and the resistance and impedance magnitude to approach a value of 400 ohms, which is representative of unbalanced broadcast antennas in this frequency range.

For automobile receivers, dummy antennas of various types have been used. If the set is intended for a particular car and antenna, a dummy representing those conditions should be devised and used. One general dummy design consists of 7 ohms in series with 160 μ mf, the latter shunted by 80,000 ohms. Another design consists of 100 μ mf in series with 20 ohms. Another is merely a 40- μ mf capacitor.

For sets with built-in loop antennas, the usual source of test signal is a fixed loop producing the test field. This is described in Sec. 10.

For the higher frequencies special cable matching the generator output circuit is desirable.²

3. *Standard Dummy Load.* This is a pure resistance equal to the 400-cycle impedance of the voice coil of the loud-speaker supplied with the receiver. The load resistor should be capable of dissipating the maximum power output of the receiver without an appreciable change in resistance. A thermal type of voltmeter is used for determining accurately the rms voltage across the load resistor.

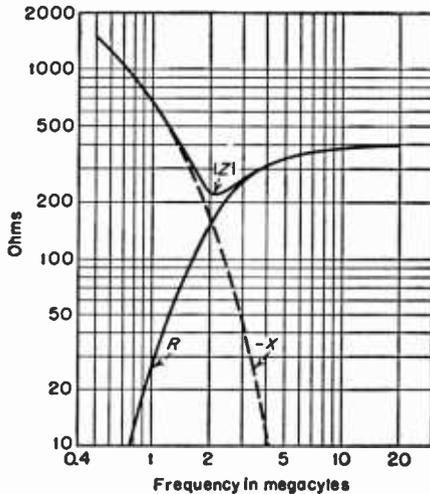
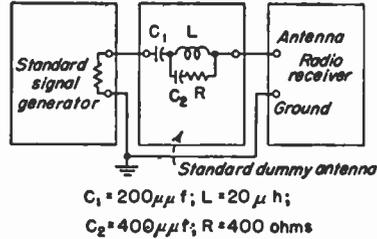


FIG. 2. Construction, use, and impedance characteristics of standard dummy antenna representing open antennas for broadcast reception in the m-f and h-f bands.

¹ Haeff, A. V., T. E. Hanley, and C. B. Smith, Wide-range Ultra-high-frequency Signal Generators, *Proc. IRE*, October, 1947, pp. 1137-1143.

² Peterson, Arnold, Output Systems of Signal Generators, *Gen. Radio Experimenter*, June, 1946.

4. *Harmonic-measuring Circuit.* For this purpose a harmonic analyzer capable of measuring frequencies up to 10,000 cycles can be used. The instrument should have sufficient selectivity to separate each harmonic and should give the amplitude of each within an accuracy of 5 per cent. However, in normal practice, test equipment is generally used by which a composite test is made evaluating the total harmonic content as a single figure.

10. *Sensitivity Test.* Sensitivity is determined by impressing an r-f voltage of 400-cycle 30 per cent modulation in series with a standard dummy antenna and adjusting the intensity of the input voltage until normal test output is obtained at resonance for various carrier frequencies.

A receiver provided with a loop antenna is tested by introducing signal inductively from a coaxial coil fed by the signal generator, or by introducing the test voltage conductively in series with the loop antenna. With the driving coil, which is the usual method, the field strength at the loop antenna is

$$E = 18.85NA^2 \frac{I}{X^3}$$

where E = equivalent electric field intensity, μv per m

N = number of turns in driving coil

A = radius of driving coil, cm

I = current in driving coil, ma

X = distance between centers of driving coil and loop antenna, m

The value of X should be at least twice the maximum dimension of either the driving coil or antenna, and much less than the distance to surrounding objects, such as the walls of the screened room.¹

Detector sensitivity is a measure of the sensitivity of that part of the receiver including the detector and the audio amplifier. It is expressed in decibels below 1 volt, or as the equivalent fraction of 1 volt, and states the input of 30 per cent 400-cycle voltage necessary for normal test output. It is often desirable to obtain the result indirectly by measuring the sensitivity on the last i-f grid and correcting this value with the separately measured gain of the last i-f stage.

11. *Selectivity Test.* Selectivity is determined by tuning to the test frequency, observing the sensitivity, and then with the receiver adjustments undisturbed, measuring the r-f input signal necessary to give normal test output at other frequencies. Points usually recorded are the discrimination at ± 10 and ± 20 kc (which are called *selectance values*) and the band width at 6, 20, and 60 db. In most cases it is not necessary to disable the a-v-c system because the level at the output of the receiver and at detector are constant, whence the a-v-c bias is constant.

12. *Electric Fidelity Test.* This is determined by tuning the receiver to 1,000 kc, maintaining the input at 5,000 μv and 30 per cent modulation, varying the modulation frequency from 30 to 10,000 cycles, and taking readings of relative output voltage at convenient modulation frequencies. The results at various a.f.s are expressed in terms of the value at 400 cycles, which is taken as zero decibels or 100 per cent.

13. *Image Rejection Test.* The ability of a superheterodyne to suppress the undesired image signal is measured as in the sensitivity test except that the signal generator is adjusted in frequency to deliver the image signal. This value of image sensitivity is then compared with the regular sensitivity to compute the image rejection, which is expressed in decibels or as a voltage ratio.

14. *I-f Rejection.* Any superheterodyne is subject to some extent to interference from a station transmitting on or near the i.f. of the set. The receiver is tuned to the r.f. of interest and the signal generator to the i.f., e.g., 455 kc. In this way the i-f sensitivity is measured. The i-f rejection is obtained by comparison with the normal sensitivity for the particular r.f.

In the case of f-m receivers, the sensitivity to i-f interference at the antenna terminals of the set is usually measured with an f-m signal of ± 22.5 -kc swing varying at

¹ Practical procedures and designs of driving coils for testing loop receivers can be found in W. O. Swinyard, *Measurement of Loop-antenna Receivers*, *Proc. IRE*, July, 1941, pp. 382-387.

a rate of 400 cps. The use of f.m. for the test signal determines the ability of the front-end circuits of the receiver to reject this spurious signal and avoids the test result depending on the behavior of the detector and the limiter, if any. Further treatment of tests on f-m receivers is given in Sec. 22.

15. Maximum Undistorted Output. This is determined by increasing the output in successive steps and measuring both the power output and the percentage of harmonics. The maximum undistorted output is the least power output which contains a total rms harmonic voltage content of 10 per cent of the fundamental voltage.

16. Maximum Output and A.V.C. For the maximum output the manual volume control of the receiver is set at maximum and a 1,000-kc signal with 30 per cent 400-cycle modulation is applied in increasing intensity from 120 to zero decibels below 1 volt (1 μ v to 1 volt). With loop receivers the observations are similar except that the intensity is in microvolts per meter and the maximum obtainable with the usual laboratory equipment is 100,000 μ v per m, or 20 db below 1 volt per m. The output using either open or loop antenna is plotted as a curve. According to latest practice, additional curves are then taken with 10 per cent and zero modulation.

To show the a.v.c., 30 per cent modulation is used and the manual volume control is reduced so that the power output for maximum signal input is $\frac{1}{2}$ to $\frac{1}{4}$ the largest value in the preceding tests. Again input signals over the full range of intensity are applied and a plot made of the observed output. A superior a-v-c system gives a curve which, with increasing signal intensity, rises rapidly at low intensities and is then relatively flat over the medium and high range. The a-v-c figure of merit is defined as the reduction below 20 db (100,000 μ v) for which the output declines only 10 db. For automobile receivers and the h-f ranges of broadcast receivers, the reference level is 46 db (5,000 μ v) instead.

17. Hum Tests. Hum originating in a-f circuits is called "residual hum" and is measured with no r-f input to the detector. The i-f amplifier is disabled by shunting $\frac{1}{2}$ μ f from the last plate to ground. Measurements are made with the volume control at its lowest and highest positions. In case there is a maximum of hum at an intermediate setting of the volume control, a measurement is made here also. The tests are normally made with an output voltmeter and may be computed and recorded in microwatts.

Hum produced by modulation of the carrier in the receiver is called "modulation hum" and is measured if it is several times as great as the residual hum. Values of modulation hum smaller than this cannot easily be measured with sufficient accuracy to be useful. For measuring modulation hum, signals at 1,000 kc of 50, 5,000, 100,000, and 2,000,000 μ v and of 400-cycle 30 per cent modulation are applied, with the volume control adjusted at each signal intensity to give approximately the maximum undistorted output. At each signal intensity the observations consist of noting the audio voltage with the 30 per cent 400-cycle modulation and then switching off the modulation and noting the voltage of each hum component. The results are converted to per cent modulation by taking the ratio of hum-component voltage to 400-cycle voltage and multiplying by 30. For example, if the voltage of a particular hum component is 0.1 volt and the 400-cycle voltage is 1.0 volt, the noise modulation for this hum component is 3 per cent.

A simplified single measurement of either residual or modulation hum is often made by observing the rms total of all hum components. In this way a single power or a single hum-modulation-percentage value is obtained.

18. Tests Involving Noise. The 1938 IRE Standards include a measure of the amount of the random types of noise originating in a receiver. This is called "e.n.s.i." standing for equivalent noise side-band input. The result is expressed in microvolts, or decibels below 1 volt, at the antenna terminal. Typical values for broadcast receivers of the home and automobile types always exceed 110 db. For home receivers operating in the m-f band, with an r-f stage and typical antenna-coupling voltage gain of 15 to 18 db, the e.n.s.i. normally amounts to 121 to 123 db. Automobile receivers with antenna voltage stepups as high as 26 db have typical e.n.s.i. of 132 to 137 db.

A home receiver without an r-f stage has increased noise because converter tubes are more noisy than amplifier tubes, and such a set may be expected to have an e.n.s.i. of about 117 db.

The significance of e.n.s.i. may be appreciated when it is noted that satisfactory reproduction of a broadcast program for the average listener requires a signal-to-noise ratio of 40 db, and mediocre reproduction requires a ratio of 20 or 25 db. For example, a set with an e.n.s.i. of 117 db requires for satisfactory reproduction a signal strength of 77 db, whereas a set with an e.n.s.i. of 137 db will give equally good reproduction on a much smaller signal, *viz.*, one of 97 db below 1 volt.

19. E.N.S.I. Test. The procedure for measuring e.n.s.i. is as follows: (1) select a convenient carrier level of about 100 db ($10 \mu\text{v}$) and call this $(E_s)_{\text{db}}$; (2) with this carrier applied without modulation, measure the rms noise voltage delivered by the receiver, calling this E_n' ; (3) switching on 400-cycle 30 per cent modulation and inserting a 400-cycle band-pass filter in the output circuit of the receiver (which will remove practically all the noise), measure the rms 400-cycle output, calling this E_s' ; (4) compute e.n.s.i. from the formula,

$$(E_n)_{\text{db}} = (E_s)_{\text{db}} + \left(\frac{E_s'}{E_n'} \right)_{\text{db}} + 10$$

Example: Consider a set measured at 100 db and giving signal and noise outputs of 50 and 10 volts, respectively. The output signal-to-noise voltage ratio is 5, or 14 db. The e.n.s.i. is then $100 + 14 + 10 = 124$ db.

If E_n and E_s are in microvolts and E_s' and E_n' in volts, the e.n.s.i. in microvolts can be obtained from the relation,

$$E_n = 0.3E_s \left(\frac{E_n'}{E_s'} \right)$$

Example: To illustrate the application of this formula, consider the example just given. The value of E_s is $10 \mu\text{v}$, corresponding to 100 db, and E_n'/E_s' is $1/50$ or 0.2, whence $E_n = 0.3 \times 10 \times 0.2 = 0.6 \mu\text{v}$, which is the same as 124 db.

E.n.s.i. is an indication of the amount of thermal and tube noise referred to the antenna terminal. It is not intended to show the amount, or the presence, of other types of disturbance, such as hum, atmospherics, or noises due to faulty operation. It is not important with insensitive sets because they have insufficient amplification to build up the thermal and shot noise to appreciable magnitude. The significance of the name "equivalent noise side-band input" is this: a single side frequency 400 cycles removed from the carrier and having this (e.n.s.i.) strength would produce an rms output of 400 cycles just equal to the rms value of the noise output produced by the receiver. The normal a-m signal has an upper side frequency and a lower one with such phase relations that in the detector the audio voltage due to the beating of the upper side frequency with the carrier adds to the audio voltage due to the beating of the lower side frequency with the carrier; *i.e.*, the total audio power of E^2/R is quadrupled in comparison with the value it would have if only one side frequency were present. However, with noise the side frequencies are random in phase, with the result that the audio output from components of higher frequency than the carrier adds, in terms of power to the output from components of lower frequency, a doubling rather than a quadrupling of the power. With noise there is an absence of cooperation between the upper and lower side bands, and this result is represented by considering the e.n.s.i. to be a single side frequency.

E.n.s.i., and especially noise figure which is presented below, have become matters of importance with receivers for the v-h-f and higher frequency ranges. One reason is the limited tube input impedances, because these limit the voltage gain which can be realized in coupling networks, with the result that the early coupling networks of the receiver cannot increase the voltage of a weak signal sufficiently to make it amply override the random noise produced in the early resistances and tubes of the receiver.

20. Noise Figure. The value of the e.n.s.i. of a receiver does not indicate directly whether the noise is large or small in comparison with an ideal receiver. For this reason, another measure of noise, called the "noise figure" or "noise factor" or "excess noise increment," has come into wide use, especially for receivers designed for the v-h-f, u-h-f, and s-h-f bands. This quantity, broadly speaking, gives a comparison between the actual noise output of the receiver and the smaller noise output which would be produced if the only source of noise were from the antenna resistance. Thus, the actual receiver is compared with an ideal having no internal sources of noise and reproducing only the noise originating in the resistance of the antenna. However, from a stricter standpoint, the noise figure is independent of the audio-amplifier characteristics and of the total gain of the receiver. It is a measure of the noisiness of the receiver resulting from the introduction of noise in the early coupling systems and tubes. (A general treatment of random noise as produced by resistances and tubes is given below.)

The noise figure is expressed in decibels or as a power ratio; it is not customary to state it as a voltage or current ratio. The larger the noise figure, the noisier the receiver. Typical values range from 8 to 30 db. The noise figure is of interest with any receiver having sufficient gain so that the ability to receive weak signals is limited only by the presence of noise originating in the receiver.

Before giving the precise definition of noise figure, it is advisable to describe *available power*. Any source of voltage having an open-circuit value E and associated with an internal resistance R will deliver maximum power to a load of matched value, *i.e.*, a load equal to R . The load voltage is then $E/2$, and the current is $E/2R$. The maximum load power is the product of this voltage and current and is, therefore, $E^2/4R$. This is called the *available power* from the particular source.

An actual or dummy antenna of resistance R can be shown to have an open-circuit rms thermal-noise voltage equal to $\sqrt{4KTRB}$, where K is a constant with the value 1.37×10^{-23} , T is the absolute temperature (usually taken as 290°), R is the resistance in ohms, and B is the total band width in cycles between the 3-db points. The available noise power is KTB . For the noise figure, all internal noise sources are referred to the input of the receiver in terms of power, and the noise figure is the ratio which the sum of this power plus KTB bears to KTB , expressed in decibels or as a power ratio. A perfect receiver reproduces only the noise of the antenna and has a noise figure expressed as a power ratio of unity, or zero decibels.¹

21. Noise-figure Test. The noise figure of a receiver is best measured with a source of random noise, such as a temperature-limited diode,² *i.e.*, one without space charge. The noise generator and dummy antenna are connected, but the noise generator is not yet turned on. An observation is made of the noise output power P_1 of the i-f amplifier under these conditions. Then turn on the noise generator and increase its output until the noise output of the i-f amplifier has increased to $2P_1$. Note the direct space current I of the diode (which is the means of adjusting the noise generator output). It can be shown that the noise figure of the receiver in decibels is

$$10 \log_{10} (20 IR)$$

where R is the resistor through which the noise current flows. For example, with 40 ma and 70 ohms, the value of $20IR$ is 56, corresponding to 17.5 db.³

E.n.s.i. and the noise figure are measures of performance with regard to noise and do

¹ For further treatment of noise figure, see H. T. Friis, *Noise Figures of Radio Receivers*, *Proc. IRE*, July, 1944, pp. 419-422; M. J. O. Strutt, "Ultra- and Extreme-short Wave Reception," pp. 67-79, *Van Nostrand*, 1947; Morris Allen, *Method of Determining Receiver Noise Figure*, *Tele-Tech.*, January, 1948, pp. 38-39, 77-78.

² Johnson, Harwick, *A Coaxial-line Diode Noise Source for U-H-F*, *RCA Rev.*, March, 1947, pp. 169-185.

³ For proof of this formula and other discussion, reference can be made to the following papers: L. A. Moxon, *Noise Factor*, Part 1—A New Conception of Receiver Sensitivity and Signal-noise Ratio, *Wireless World*, December, 1946, pp. 391-393; L. A. Moxon, *Noise Factor*, Part 2—Methods of Measurement and Sources of Test Signals, *Wireless World*, January, 1947, pp. 11-14; and J. M. Pettit, *Specification and Measurement of Receiver Sensitivity at the Higher Frequencies*, *Proc. IRE*, March, 1947, pp. 302-306.

not give information as to receiver sensitivity. A need for a single quantity indicating performance with regard to both noise and sensitivity has been recognized for more than a decade in connection with military and naval receivers, and has led to the specification that the volume control of the receiver be reduced if necessary to avoid output noise exceeding a stated amount. The sensitivity of the set was then measured with this position of the volume control, and the result required to pass the sensitivity provision of the specifications. For example, in one of the armed services a receiver for headphone reception of voice was tested by adjusting the volume control to give an output of 600 μw of noise with carrier present but unmodulated. Upon application of standard 30 per cent 400-cycle modulation, the output was required to increase to 6 mw, corresponding to a signal-to-noise-ratio of approximately 10 db. Some successive approximation in the test in this case is necessary until the proper volume-control setting and carrier level are found. This carrier level is then the measured sensitivity.

A slightly different procedure for obtaining a sensitivity measurement which takes noise into account has been proposed by J. M. Pettit.¹ He suggests that the quantity be called the *combined sensitivity figure*. Successive approximation is avoided by arbitrarily establishing in advance the standard noise output. With the signal generator and dummy antenna connected but the generator not operating, the volume control of the receiver is advanced until standard noise output is delivered. This setting of the volume control is thereafter not disturbed and is called the *standard gain setting*. The signal generator is then started and a regular sensitivity measurement made. This value is the combined sensitivity figure.

22. Testing F-m Broadcast Receivers.² The standard modulation is 30 per cent of the full 75-kc swing to either side, or ± 22.5 kc. The a.f. remains at 400 cycles. The least value of carrier to give normal test output, which is the same as with a.m., is called the *maximum-sensitivity-test input*.

Another test is made with the full 75-kc deviation. Starting with a very weak signal, the input signal level is increased and the volume control reduced, maintaining normal test output, until the output distortion, measured as the rms sum of all harmonic voltages, declines to 10 per cent. This is the *maximum undistorted output*, and the signal input at this point is called the *maximum-deviation sensitivity-test input*.

The *quieting-signal sensitivity-test input* is the input necessary to give a 30-db signal-to-noise ratio upon the application of the standard 400-cycle 22.5-kc deviation modulation. The receiver is adjusted for greatest sensitivity except that the volume control may be reduced to prevent audio overload.

These three values of sensitivity are expressed in terms of input signal. It is usually considered that the poorest one of the three determines the usefulness of the receiver.

A fourth test determines the *deviation-sensitivity input*, which indicates the amount of deviation necessary for maximum undistorted output at standard mean-signal input. The result of this test is expressed in kilocycles deviation, e.g., 5 kc, indicating a swing of this much in each direction. The test shows whether the useful sensitivity of the receiver is limited by the audio gain. The standard mean-signal input is an available power of 90 db below 1 watt, or 1,100 μv acting through a 300-ohm dummy antenna.

Observations with various dummy antennas have shown the most reliable measurements on balanced-input receivers to be obtained with a signal generator having balanced-output connections and a dummy antenna consisting of 600-ohm resistors connected to each high lead of the generator, the other ends of these resistors being joined through a 400-ohm resistor. The receiver is connected across the 400-ohm resistor, and the grounds of the signal generator and receiver are connected. Each of the 600-ohm units includes the internal resistance of one side of the signal generator, so that they must actually be somewhat less to give a 600-ohm total. If the common ground connection between the generator and receiver chassis is removed, good results are still obtained.

¹ *Loc. cit.*

² IRE, "Standards on Radio Receivers: Methods of Testing Frequency-modulation Broadcast Receivers: 1947."

A simpler dummy antenna for a balanced generator consists only of two 150-ohm resistors, including the generator resistance. This gives good results if the grounds of the generator and receiver are *not* connected.

For unbalanced generators, satisfactory results are obtainable with two 150-ohm resistors, reduced as necessary to give a total of 300 ohms for the generator and the

Average Characteristics of Home Socket-power Receivers

Characteristics*	Receivers†				
	Four tubes plus rectifier, a-c/d-c, two-gang	Five or more tubes (avg., 5.4) plus rectifier, a-c/d-c, two-gang	Four tubes plus rectifier, a-c, two-gang	Five or more tubes (avg., 5.5) plus rectifier, a-c, two-gang	Avg 6.0 tubes plus rectifier, a-c, three-gang
With test loop:					
Sensitivity, db/m \ddagger	75	83	80	81	92
E.n.s.i., db/m \ddagger	95	101	105	103	106
Band width at 6 db, kc.....	7.1	7.7	7.0	7.3	7.2
Band width at 40 db, kc.....	31	31	28	28	24
Image rejection, db.....	37	41	38	41	> 74
I-f rejection, db.....	37	39	38	34	67
Loop figure of merit, db.....	-7.7	-7.1	-1.4	-4.2	-2.0
Loop band width at 6 db, kc.....	24	23	21	23	25
With dummy antenna:					
Sensitivity, db.....	93	101	90	97	108
E.n.s.i., db.....	117	119	120	121	126
Band width at 6 db, kc.....	7.9	7.7	6.8	7.5	6.8
Band width at 60 db, kc.....	54	50	47	44	37
Image rejection, db.....	32	34	34	38	79
I-f rejection, db.....	40	40	40	35	69
Voltage gain in antenna coupling, db.....	12	11	14	13	14
Band width of antenna coupling at 6 db, kc.....	27	29	20	27	25
Receiver characteristics:					
A-v-c figure of merit, db.....	41	45	36	38	49
Max undistorted power output, watts.....	0.9	1.0	2.3	2.2	2.5
Conversion gain, db.....	30	30	30	34	30
Gain of i-f stage at 455 kc, db.....	39	39	42	40	38
Band width of i-f stage at 6 db, kc.....	19	17	15	15	14
Band width of i-f stage at 20 db, kc.....	48	44	40	38	39
Detector sensitivity, db.....	12	10	6.7	4.5	2.1
Power consumption, watts.....	28	33	60	68	72
Total B drain, ma.....	62	68	62	85	77

* Frequency = 1 Mc. Test output = 0.5 watt.

† A-c/d-c sets operated on a.c.

‡ Values are in decibels below 1 volt per m.

two resistors in series. It is necessary to avoid a direct connection between the grounds of the two chassis.

Two precautions have been suggested¹ in this regard: (1) the polarity of the generator output should be reversed, and (2) the power-line connection for either the generator or the receiver should be moved to a different outlet. No change in the measured sensitivity of the receiver should result from either of these two changes.

23. Typical Performance. Averages of a large number of preproduction and manufactured broadcast receivers which were tested in the Hazeltine Laboratories during the first two years after the Second World War show the performance indicated in the

¹ Rankin, John A., Receiver Input Connections for U-h-f Measurements, *RCA Rev.*, April, 1942, pp. 473-481. This gives data at 43 Mc.

table on page 785. The observations marked "With test loop" show the performance with the loop antenna which is generally a part of broadcast receivers. The sensitivity in this case is given as the field strength necessary to produce an output of 0.5 watt. This is stated in decibels below 1 volt per m, and the e.n.s.i. (as described in Sec. 19) is stated in the same units.

The loop figure of merit, as described in Sec. 32, is stated in decibels. The fact that this is negative shows that a field strength of, for example, 1,000 μv per m produces less than 1,000 μv on the first grid.

The a-v-c figure of merit, as described in Sec. 79, is the change in input signal intensity necessary to produce a 10-db change in output. Values of this quantity are seen to range from 36 to 49 db.

The maximum undistorted power output, which is defined in Sec. 15, is the largest output which will not produce more than 10 per cent rms sum of harmonic voltages, corresponding to 1 per cent total harmonic power.

The detector sensitivity is the number of decibels below 1 volt characterizing the signal that is just sufficient to produce 0.5 watt output. It is seen that this varies from 12 to 2.1 db, corresponding to 0.25 to 0.8 volt.

NOISE IN RECEIVERS

Any one of numerous conditions can produce disturbance to radio reception. The cause may be external to the receiver, such as atmospheric, cosmic static, interference from unfiltered electrical appliances like elevators and shavers, ignition noise from automobiles, or interference from diathermy equipment. Internal noise can arise in the form of hum, hiss originating in early circuits and tubes of the set, and clicks or other disturbance from faulty connections, tubes, fixed capacitors, or other defective parts.

24. Necessary Signal-to-noise Ratio. Signal and noise can be measured and compared in terms of rms voltage or current, or in terms of power, and the ratio of signal to noise be expressed in decibels. For satisfactory reception of the typical broadcast program, a signal-to-noise ratio of 40 db is necessary. A value of 30 db has been established as determining the "interference test output" in certain standard receiver-test procedures. In television 30 db is about the minimum signal-to-noise ratio for a satisfactory picture. In radiotelephone service, where intelligibility is the chief criterion, 18 db is sufficient. Other services, such as radio direction finding, where the indications may be read on meters, will operate with much lower signal-to-noise ratios and may even function with less signal than noise corresponding to a signal-to-noise ratio expressed by a negative number of decibels.

25. Sources of Noise External to Receiver.¹ Careful measurements by personnel of the Bell Telephone Laboratories using directional antennas have shown that noise of extraterritorial origin and appreciable amount is received. At 18 Mc using a rhombic antenna K. G. Jansky² observed noise values from 11 to 21 db above the amount available from thermal agitation in the resistance of the antenna. The noise came predominantly from the direction of the denser part of the Milky Way. At three frequencies in the s-h-f band between 3,000 and 30,000 Mc, G. C. Southworth³ measured noise from the sun. The amount at the lower frequencies of this range was in good agreement with a theory worked out on the basis of black-body radiation of long heat waves and short radio waves.⁴

In the m-f and lower bands terrestrial static is a common disturbance. To reduce

¹ Thomas, H. A., and R. E. Burgess, "Radio Research Special Report No. 15—A Survey of Existing Information and Data on Radio Noise over the Frequency Range 1-30 Mc/s," H. M. Stationery Office, London; also available from British Information Services, New York. This is a pamphlet of 126 pages and is the best available single source of data on the subject.

² Jansky, K. G., Minimum Noise Levels Obtained on Short-wave Radio Receiving Systems, *Proc. IRE*, December, 1937, pp. 1517-1530.

³ Southworth, G. C., Microwave Radiation from the Sun, *J. Franklin Inst.*, April, 1945, pp. 285-297.

⁴ A recent discussion of extraterritorial noise is given in Sec. III of K. A. Norton and A. C. Omberg, The Maximum Range of a Radar Set, *Proc. IRE*, January, 1947, pp. 4-24.

its effects use is made of (1) directional antennas, (2) narrowing of frequency channel as much as the signal will allow, (3) special circuits which more or less disable the receiver for very short lengths of time while a noise pulse is arriving, and (4) other special equipment and methods.

In the category of man-made interference, diathermy has been the worst disturbance in the h-f band. In the paper by Jansky referred to, he reported that during the day it would at times take hours, with a receiver having a 1.6-kc band width, to locate in a band of 1.3 Mc width centered at about 18 Mc any point where observations for cosmic static could be made free of diathermy interference.

Interference from electrical appliances is well known. A British standard BSS. No. 800-1939 designates 500 μv as the maximum radio output of an appliance regarded as "radio-interference-free." If the attenuation from the appliance to the receiver has the typical value of 25 db and a 40-db ratio of full-modulation signal to the noise is required, this results in the necessity for a signal of 2,800 μv , or 51 db below 1 volt at the receiver input terminals. It is clear that most of the sensitivity range of the receiver is useless under these circumstances.

The usual methods of curing interference due to appliances¹ are (1) to connect a low-pass filter in the supply line as close as practicable to the appliance, so as to prevent the line from acting as a conductor or antenna for the interference, (2) to connect capacitors at sparking elements of appliances such as motor brushes, and (3) to locate the antenna where the noise is not encountered and use a leadin having a balanced arrangement or shielding, or both.

Automobile ignition systems produce interference in the h-f, v-h-f, and u-h-f band. Measurements with a receiving antenna of 35 ft in height at a distance of 100 ft showed field strengths of roughly 40 μv per m for the horizontally polarized component and 50 μv per m for the vertical at various frequencies from 40 to 450 Mc. These results were reported by R. W. George,² who also gave theoretical curves which can be used to get field strength at distances greater and less than the 100 ft for which the measurements were made.

Ignition interference is an example of *impulse noise*, which is to be distinguished from fluctuation noise. Impulse noise consists of discrete separate pulses with relatively long quiet intervals between. Each pulse shock-excites the usual radio tuning circuits. The resulting oscillations have the frequency of the various tuned circuits and die out exponentially; *i.e.*, each disturbing pulse produces a damped train of oscillations. Ordinarily, this train dies away considerably before the occurrence of the next pulse. Fluctuation noise, on the other hand, is always present with an instantaneous amplitude determined by statistical probability laws. Examples of it are resistance noise and tube noise, mentioned below.

26. Noise Sources in the Receiver. After noise due to poor contacts, defective parts, and microphonics³ has been removed from a receiver, it will, with the usual a-c power supply, still deliver noise in the form of *hum*.⁴ By good design and maintenance work, this can be reduced until it is no longer appreciable. Whenever hum is encountered, special attention should be paid to the low-level audio stages (where the trouble may be heater-cathode leakage⁵), the possibility of inadequate plate-supply filtering, or magnetic induction from stray fields. Another type of this disturbance is "modulation hum," which is present only while a carrier is being received. This is produced by nonlinearity in the r-f and i-f stages plus a hum voltage, so that the hum modulates

¹ Smith, W. W., *et al.*, "Radio Noise Reduction Handbook," 2d ed., Radio, Ltd., Santa Barbara, Calif., 1940. "Practical Radio Noise Suppression," Sec. 10 in "MYE Technical Manual," pp. 310-324, P. R. Mallory & Co., Indianapolis, 1942.

² Field Strength of Motorcar Ignition between 40 and 450 Megacycles, *Proc. IRE*, September, 1940, pp. 409-412.

³ For pointers on locating defective noise-producing parts see Tracing Radio Noise, *Radio and Television Retaining*, May, 1945, pp. 66-68. Receiver Microphonics Caused by Heater-cathode Capacitance Variation, *RCA Application Note* 123, RCA Tube Department, Oct. 15, 1947. Discusses the oscillator tube.

⁴ Glaser, Mark, Killing the Hum-bug, *Service*, January, 1941, pp. 5-6, 29.

⁵ Heater-cathode Leakage as a Source of Hum, *Electronics*, February, 1940, p. 48.

the carrier. Such hum would, of course, be reproduced by a perfect detector and audio system and must be corrected by removing the cause in the r-f or i-f amplifier.

Other noise sources in the receiver include those of the random-fluctuation type producing a characteristic hisslike sound in reproducers and "grass" on oscilloscopes. These sources include resistance noise and tube noise, important in high-gain receivers and in receivers for the v-h-f and higher frequency ranges.

27. Fluctuation Noise. The irregular motion of electrons in any resistor and in the space current of any tube gives rise to the fluctuation type of noise.¹ These random motions result in small fluctuations of voltage and current. The instantaneous value of this voltage or current obeys the probability law, under which there is a certain probability that the value will fall in any range of interest. For example, there is a very small probability that a very large value will occur, so that the usual concept of peak amplitude does not apply. Instead, the probability of a large amplitude increases as the band width of the equipment is increased. There is no definite wave form. However, there is a definite rms value and also a definite average value for all positive or all negative amplitudes. These two averages are equal and each amounts to 80 per cent of the rms value.

The instantaneous amplitude in one direction or the other exceeds the rms value 32 per cent of the time, but it exceeds four times the rms value for only 6.4/100,000 of the time.

Another characteristic of fluctuation noise, before it has gone through any selective circuits, is that it has a uniform frequency distribution; *i.e.*, a 1,000-cycle band selected in the audio region has as much rms voltage as a 1,000-cycle band at 1, 10, or 100 Mc.

Fluctuation noise establishes a limit to the amount of amplification that can be usefully employed. At the input to a receiver or amplifier, the signal must be large enough so that it will not be lost in the noise associated with resistances and tubes in the low-power portion of the equipment.

28. Thermal-agitation Noise. The random motion of electrons, under the molecular agitation due to temperature, produces noise across the terminals of any resistor. A parallel resonant circuit or other passive network manifests the same effect to the extent corresponding to the resistive component of the impedance. The magnitude of the open-circuit rms voltage is given by the relation

$$E = \sqrt{4KTRB}$$

where K = Boltzmann constant = 1.37×10^{-23}

T = absolute temperature, °K (about 300 for room temperature)

R = resistance, ohms

B = band width, cps

For radio purposes it is more convenient to express this in microvolts, ohms, and megacycles. Making these changes and inserting the numerical values of K and T , we obtain

$$E = 0.13 \sqrt{RB}$$

Values of E for wide ranges of resistance and band width are plotted in Fig. 3.

For the usual noise computations, which have a limited accuracy, the determination of B will be easy. For a selectivity curve having a shape like that of a single tuned circuit, the value of B is the total frequency interval between the two points where the curve is 3 db down. For more involved cases, a graphical determination can be made by using graph paper with linear graduations in both directions and plotting frequency as abscissas and power (or voltage squared or current squared) as ordinates. Measure the area under the curve and determine the width of a rectangle of the same area having a height equal the highest single ordinate. The width of this equivalent rectangle is the desired value of B .

¹ Landon, V. D., The Distribution of Amplitude with Time in Fluctuation Noise, *Proc. IRE*, February, 1941, pp. 50-55. Pierce, J. R., Noise in Resistances and Electron Streams, *Bell System Tech. J.*, January, 1948, pp. 158-174.

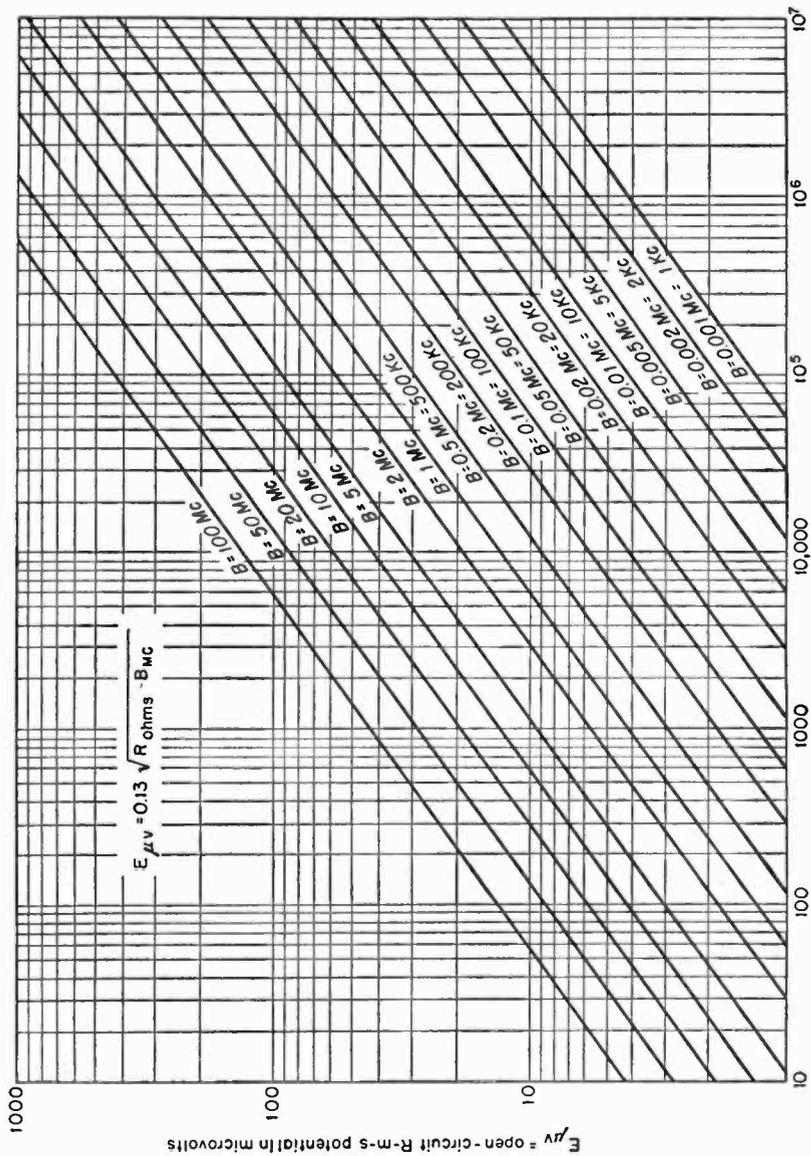


Fig. 3. Thermal noise for various resistances and band widths.

An important quantity associated with the thermal noise voltage is the *available power* from such a source. As with any source of open-circuit voltage E acting through an internal resistance R , the maximum power which can be delivered, defined as the available power, is obtained by matching the load resistor to the internal resistance. The value of the available power is $E^2/4R$. Applying this to the case in hand, the available thermal-noise power is $(0.13)^2RB/4R$, or $W = 0.004B$. Here B is in megacycles, and W is in microwatts. The following table gives values computed from this formula:

Band width, B , Mc	Available power, W , μw	Decibels below 1 watt
10^{-4} (100 cycles)	4×10^{-7}	184
10^{-3} (1,000 cycles)	4×10^{-6}	174
10^{-2} (10 kc)	4×10^{-5}	164
10^{-1} (100 kc)	4×10^{-4}	154
1	4×10^{-3}	144
10	0.04	134

An alternative expression for available noise power is 4×10^{-21} watt per cycle of band width.

29. Tube Noise. Fluctuation noise originates in tubes due chiefly to (1) the random nature of emission at the cathode and (2) the irregular and changing distribution of current between plate and screen if the tube has a screen. The first of these was discussed by Schottky¹ in 1918, along

with resistance noise, all under the name *schroetteffekt* (German for "shot effect") from the similarity of the sound to that of many small pellets falling on a hard surface. During the years this term has come to be restricted by most workers to tube noise resulting from random cathode emission so that resistance noise is not included. The name "partition noise" is often applied to tube noise resulting from the irregularities in distribution of current between plate and screen.

A common method of stating tube noise is to specify an equivalent resistance which, if connected at the grid and if the tube itself were noiseless, would produce the same noise output due to the room-temperature thermal noise in the resistance. This is more convenient than a voltage statement because the resistance value does not depend on band width while the voltage value does.

Certain precautions in the use of equivalent grid resistance to represent tube noise should be borne in mind, *viz.*, that the concept does not apply if the tube has feedback nor if it has input loading such as that due to transit-time loss. To avoid these limitations, it has been suggested that the internal cathode resistance of the tube be considered as the source of the noise. This is discussed later.

The noisiest tube condition is with temperature saturation, *i.e.*, no space charge. In this case all electrons emitted are immediately drawn to the plate. A diode operated in this way is, in fact, the best standard source of noise for use in measuring the noise figure of a receiver. The presence of the space charge, which makes amplification possible, greatly reduces the tube noise, which is a very fortunate effect.

Values of equivalent grid resistance expressing the noisiness of various tubes have magnitudes² generally lying in the following ranges:

Tube and circuit	Ohms	Tube and circuit	Ohms
Triode amplifiers.....	200-2,300	Gradual-cutoff pentode amplifiers	2,400- 14,000
Triode mixers.....	900-6,000	Pentode mixers.....	2,800- 35,000
Sharp-cutoff pentode amplifiers..	700-7,000	Hexode and heptode mixers.....	190,000-300,000

Mixer noise increases beyond these values if the oscillator injection is insufficient. For amplifiers the effective input noise increases, but the output noise decreases as the gain is reduced by the application of more negative grid bias.

¹ Schottky, W., Spontaneous Current Fluctuations in Various Conductors, *Ann. Physik*, **57**, 541-567, Dec. 20, 1918.

² See Harris, W. A., Fluctuations in Vacuum Tube Amplifiers and Input Systems, *RCA Rev.*, April, 1941, pp. 505-524, which includes applications of extensive theoretical and practical work done with B. J. Thompson and D. O. North.

Formulas for equivalent grid resistance expressing tube noise in terms of the usual currents and conductances are available in the literature.¹

To avoid the limitations of the grid-resistance concept of tube noise, which are mentioned above, it has been proposed by Harold A. Wheeler that the internal space-charge resistance in series with the cathode be considered as the source of the noise. In this case the internal cathode resistance has a temperature of 0.6 of the absolute temperature of the heated cathode. The temperature value, after application of the 0.6 factor, is approximately twice the absolute room temperature for oxide-coated cathodes, three times for thoriated filaments and five times for pure tungsten filaments.

For triodes this cathode resistance is $1/(g_m + g_p)$, where g_m is the transconductance and g_p the plate conductance or reciprocal of the usual a-c plate resistance. The following values for several high-transconductance triodes have been computed for normal operation with the highest recommended plate voltages:

Tube type	Transconductance, g_m , μmhos	Plate resistance, $r_p = \frac{1}{g_p}$, ohms	Cathode noise resistance, ohms
6J4	12,000	4,500	82
6K4	5,500	3,650	173
6J6 (per triode)	5,300	7,100	184
6C4	2,200	7,700	430
8002	2,200	11,400	440
3A5 (per triode)	1,800	8,300	520

For pentodes and other multiclement tubes the cathode noise resistance is the a-c resistance which would be measured from the cathode to all the other electrodes joined by capacitors while at their normal d-c operating potentials. The noise associated with this resistance is a minimum value which is increased by partition noise. These incremental sources of noise are considered as due to the cathode resistance being at a higher temperature than the initial 0.6 value.

Sources of fluctuation noise in the first circuits and tubes of a receiver are usually all that have to be considered because it is only these that are followed by sufficient amplification to deliver appreciable noise at the output.

Calculations of various noise effects in the early circuits of receivers² can show much of the quantitative actions.

As general design and developmental guides, the following rules for reducing noise effects can be given:

1. Use as much antenna directivity as possible.
2. Endeavor to have a large enough antenna to intercept sufficient signal to predominate over noise sources in the receiver.
3. Reduce the band width of the receiver as much as the signal characteristics permit.
4. Have as high voltage gain as possible in the coupling between the antenna and the first tube. For this purpose select as the first tube one with as high input resistance as possible.
5. Use an r-f stage to avoid the large noise of mixers early in the circuit.
6. Consider the use of triodes rather than pentodes. In the v-h-f and higher bands consider the grounded-grid triode. At u-h-f and higher bands consider crystal rather than tube mixers.
7. Select particular types of tubes for low noise, especially for the first one or two sockets of the receiver.
8. Use preemphasis in transmission and deemphasis in reception of low-energy portions of band of modulation frequencies.
9. Use volume compression in transmission and volume expansion in reception.

¹ *Ibid.*

² Stolze, Wm. J., Input Circuit Calculations for F-m and Television Receivers, *Communications*, February, 1947, pp. 12, *et seq.* Wallman, Henry, Alan B. MacNee, and C. P. Gadsden, A Low-noise Amplifier, *Proc. IRE*, June, 1948, pp. 700-708.

10. Use f.m. or pulse modulation rather than a.m.
 11. Consider the use of a squelch circuit. If impulse noise is encountered, consider noise limiters. Squelch- and noise-limiting circuits are discussed in Secs. 93 and 94.

30. External Cross Modulation. A form of interference to which all types of receivers are subject can exist where two or more stations lay down very strong signals in a given area. The trouble arises from poor contacts between various extended conductors in the vicinity, such contacts acting as frequency converters and thereby producing new r-f signals which reach nearby receivers either conductively or by radiation. For example, in Seattle about 1938, three stations on 920, 970, and 1,270 kc each produced strong signals in a large residential area. All three stations together were heard at 620 kc (which is $920 + 970 - 1270$) as well as at 1,220 kc ($920 + 1270 - 970$) and at 1,320 kc ($970 + 1270 - 920$). At three additional frequencies, two of the three stations were heard; e.g., the 920 and 970 stations were heard at 870 kc ($2 \times 920 - 970$).

An investigation by RMA led to the suggestion that a limitation in the field strengths of overlapping coverages is advisable. A quantity called the "field product" was defined and a maximum of 0.001 recommended for it. (Values of 0.0015 to 0.005 were found in the Seattle cases where trouble existed.) Where three stations are involved, the field product is $\frac{3}{2} F_1 F_2 F_3$, where the F 's represent the field strengths of the three stations in volts per meter. If only two stations are involved, there are two field products, namely $\frac{3}{4} F_1^2 F_2$ and $\frac{3}{4} F_1 F_2^2$, and the same maximum of 0.001 applies to these.

Conductors which may form contacts and give rise to external cross modulation include house electric wiring, BX-cable sheath, antenna and ground wires, water pipes, metal lath, gas pipe, and rain spouts. The remedy is to bond the conductors together or separate them with good insulation. Whenever external cross modulation is suspected, the antenna and ground circuits of the receiver should first be checked, and then a filter tried where the power plug is inserted in the receptacle. If necessary, a new antenna with shielded lead-in can be tried, and it may be helpful to install a balanced antenna input circuit.

Accounts of theory and practical observations of cross modulation have been published.¹

LOOP ANTENNAS IN BROADCAST RECEIVERS

The convenience of the loop antenna, in comparison with outdoor or indoor open antennas, has brought it into general use for standard-band broadcast reception. The directional properties of the loop may occasionally require the user to make a slight rotation of the entire set, since the loop is usually rigidly mounted. This has not been found objectionable.

31. Voltage Induced in Loop by Wave. The passage of a wave in the direction toward which a loop is turned induces in the loop a voltage given by the formula

$$E = 0.021eNaf$$

where E = induced voltage, μv
 e = field strength, μv per m
 N = number of turns in loop
 A = area of loop, sq m
 f = frequency, Mc

The ratio E/e is by definition the series voltage in the loop divided by the field strength and is often called the "effective height" from its similarity to the effective height of an open antenna. The formula for the effective height of a loop antenna is, therefore,

$$\text{Eff. height} = \frac{E}{e} = 0.021Naf$$

¹ Ebel, A. J., A Note on the Sources of Spurious Radiations in the Field of Two Strong Signals, *Proc. IRE*, February, 1942, pp. 81-84. Eastman, A. V., and Lawrence C. F. Horle, The Generation of Spurious Signals by Nonlinearity of the Transmission Path, *Proc. IRE*, October, 1940, pp. 438-443. Foster, Dudley E., A New Form of Interference—External Cross Modulation, *RCA Rev.*, April, 1937, pp. 18-25.

where the effective height is in meters, and the other quantities are the same as in the preceding paragraph.

The effective heights of loop antennas are small in comparison with open antennas, typical loop values for 1 Mc lying in the range from 0.003 to 0.03 m.

32. Figure of Merit of Loop. The important action performed by a loop and the coupling elements, if any, between it and the first tube is to deliver voltage to the grid of this tube. The ratio of this voltage to the field strength is usually called the "figure of merit," but is also known as the "effectiveness" and the "pickup factor." This is the best single quantity for indicating the performance of a loop and includes the effect of coupling elements, if any, such as a loading coil or transformer. In terms of Q the figure of merit is the product of Q and the effective height; therefore,

$$\text{Figure of merit} = 0.021NAfQ$$

The value of the Q of a loop depends greatly on whether substantial losses are introduced by close location of the loop to the metal chassis. Typical effective values of Q for loops, with coupling circuits if any, lie in the range from 40 to 200.

Typical values of the figure of merit, being the product of effective height and Q , range from 0.12 to 6, which is -18 to $+16$ db.

33. High-impedance Loops. In many receivers the loop constitutes the entire tuning inductance of the first circuit, the only other element present being the variable tuning capacitor. The loop and the section of the gang capacitor are connected in parallel and deliver signal to the first grid. A loop of this kind is called a "high-impedance loop" in distinction to other loops of smaller inductance which furnish only part of the circuit inductance.

Measurements on 30 typical high-impedance loops designed for broadcast-band use in regularly manufactured receivers were made shortly before the war by William O. Swinyard of the Hazeltine Chicago laboratory.¹ The 30 loops gave the following data:

Arbitrary size of grouping	No. of loops	Average area		Average No. of turns	Average dist. capacitance, $\mu\mu\text{f}$	Eff. height at 1 Mc, m
		Sq in.	Sq m			
Small.....	7	29	0.019	35	7.4	0.014
Medium.....	11	42	0.027	28	11	0.016
Large.....	12	76	0.049	22	13	0.023

34. Low-impedance Loops. Under the practical conditions of small crowded cabinets, particularly with table models, it may be better to use a loop of only a few turns, as large and well placed as possible, and employ a loading coil or a transformer, along with the variable capacitor, to complete the input circuit.

Typical low-impedance loops range around 20 μh , have about five turns with a spacing of $\frac{3}{8}$ in. between turns, are made of No. 18 solid wire, and have a Q (for the loop only) of 135 at 1 Mc.

The use of a low-impedance loop with a loading coil gives rise to the possibility of having a fixed tuning capacitor and selecting stations by permeability tuning of the loading coil. This and other coupling systems for low-impedance loops are discussed in Sec. 37.

H-f Loops. A loop of one or two turns of heavy wire or copper tubing can be used for the reception of international broadcast stations. It is placed horizontally to avoid coupling with the m-f loop, which is always vertical.

Another arrangement for the h-f band is to use the m-f loop as a "plate" or capaci-

¹ Measurement of Loop Antenna Receivers, *Proc. IRE*, July, 1941, pp. 382-387. For additional data on broadcast-receiver loops, see E. M. Kendell, Receiver Loop Antenna Design Factors, *Communications*, November, 1945, pp. 62, 64.

tive element. It is used with a transformer designed for a low-capacitance open antenna.

35. Interference Rejection by Means of a Loop. Loops for m-f reception are always vertically mounted because noise interference is greater with horizontal mounting. A further improvement can be made with an electrostatic shield. Such a shield stops lines of electric force that otherwise would reach the antenna, but will carry little or no circulating current so that the magnetic field is not affected. One construction is to have numerous parallel wires each joined in comblike fashion to a common lead at one end and open at the other end.

Since interference from electrical appliances in the vicinity of the receiver is propagated with a predominant electric component, a shielded loop will reduce the effect of such disturbances. Receivers with this feature have appeared on the market in past years in a limited number of models.¹

Further rejection of noise can be obtained by means of a balanced loop, as in direction-finder practice, and this is used in some high-grade receivers.

36. Couplings for High-impedance Loops. The most common type of antenna now used for broadcast reception on the standard m-f band is the loop, so couplings for it are first discussed below, followed by couplings for sets using only open antennas.

For loop reception, no coupling circuit except the variable tuning capacitor is needed with the high-impedance loop. For radio frequencies, the loop, the capacitor, and the input circuit of the first tube are in parallel. The figure of merit of the loop, described in Sec. 32, is the voltage gain of the coupling. The Q determines the selectivity from the following relations:

$$W_6 = \frac{\sqrt{3}f}{Q}$$

$$W_{10} = \frac{3f}{Q}$$

$$W_{20} = \frac{10f}{Q}$$

where W_6 , for example, is the band width between the points at which the curve is 6 db down. Any frequency units can be used but must be the same for W and f .

With high-impedance loops there remains the requirement of providing for the connection of an outside antenna when the receiver is used in a region of low field strengths.

The most common practice is to connect the antenna to one or two turns of wire around the loop, or the electrical equivalent in the form of an autotransformer obtained by tapping the loop at one or two turns from the low-potential end. However, this type of coupling has the limitation that the voltage gain varies considerably over the band. In a typical case the voltage gain was 9 db (2.8 times) at 550 kc, 21 db (11.5 times) at 1,000 kc, and 24 db (16 times) at 1,500 kc. Such gain in the middle- and upper-frequency parts of the band gives poor image rejection and may cause cross modulation, objectionable tweets, and distortion. In the particular case the coupling system had an image rejection of about 34 db at the l-f end of the band and about 27 db at the h-f end.

One simple step to improve the performance is to insert a resistance of about 500 ohms directly in the antenna connection. One such circuit had a voltage gain of 12 db (four times) at 1,000 kc.

A more thorough remedy is to use a high-inductance primary, so as to resonate the primary circuit below the broadcast band. Windings of 30 to 55 ohms d-c resistance shunted with 50 to 300 $\mu\mu\text{f}$ have been used for this purpose.

¹ The design and explanation of a rotatable shielded loop are given by Stanford Goldman, A Shielded Loop for Noise Reduction in Broadcast Reception, *Electronics*, October, 1938, pp. 20-22. The effectiveness of such shielding can be checked by use of a ground plane and a vertical antenna connected to the signal generator. Details are given in Dudley E. Foster and C. W. Finnigan, A Method of Measuring the Effectiveness of Electrostatic Loop Shielding, *Proc. IRE*, June, 1943, pp. 253-255.

An economical and effective solution is to retain the low-inductance primary of one or two turns around the loop and add a loading coil and a damping resistor. A suitable circuit is shown in Fig. 4. The coil can be a low-cost solid-wire type. This circuit gives primary resonance either in or near the l-f end of the band, the particular point not being important because the damping by the resistor removes any difficulties in the unicontrol alignment with other tuning circuits. The following performance was obtained with this circuit when using the standard IRE dummy antenna:

Frequency, kc	Voltage gain, μ	Db	I-f rejection (455 kc), db	Image rejection, db
550	4.0	12.1	28	59
600	4.2	12.5	32	60
800	4.2	12.5	..	55
1,000	3.8	11.6	36	53
1,200	3.1	9.8	..	50
1,400	2.6	8.3	33	50
1,500	2.2	6.9	..	47

The i-f rejection in this design can be improved at the expense of the image rejection, if desired, by reducing the 275- μ h inductance of the loading coil.

37. Couplings for Low-impedance Loops. Low-impedance loops are used because the few turns required can be disposed a reasonable distance away from the chassis in crowded cabinets. The effective Q of the coupling system can then be maintained by the benefits of this spacing plus the use of an efficient powdered-iron-core loading coil or transformer.

In the loading-coil design, the loop is at the low-potential side of the circuit and the loading coil at the high side. The two are in series and are tuned by the antenna section of the variable capacitor gang. In a typical design, a 22- μ h loop made of 6 turns of No. 18 AWG wire had a Q of 147 at 1 Mc. With an iron-core loading coil of 205 μ h having 85 turns of stranded wire and a Q of 190 at 1 Mc, a circuit figure of merit of -1.5 db (0.71 voltage ratio) was obtained. The band width W_s was 15 kc.

This design was adapted to an open antenna by tapping at the junction between the loop and loading coil and connecting a 2,200-ohm resistor to the antenna terminal. With the standard dummy antenna this gave voltage gains between 9 and 12.5 db (2.8 to 4.2 times) over the broadcast band, with 2.4 kc as the value of W_s at 1 Mc.

In an alternative design, using a transformer with tuned secondary to couple the low-impedance loop to the first tube, a powdered-iron core should be used in the transformer to make the coupling and the transformer Q as high as possible. Analyses of this circuit have been given.¹

Permeability Tuning. The short supply of gang tuning capacitors in the early postwar period led to a wide use of permeability tuning in home receivers. In auto-

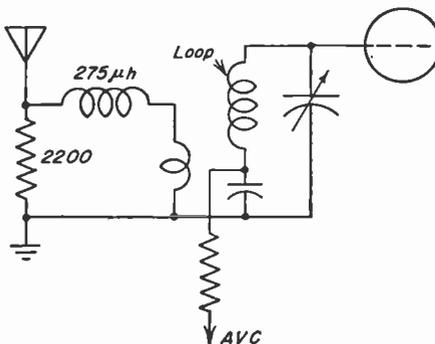


FIG. 4. Coupling for use of outside antenna with high-impedance loop.

¹ Bachman, W. S., Loop Antenna Coupling-transformer Design, *Proc. IRE*, December, 1945, pp. 865-867. Vladimir, L. O., Low Impedance Loop Antenna for Broadcast Receivers, *Electronics*, September, 1946, pp. 100-103. Browder, J. E., and V. J. Young, Design Values for Loop-antenna Input Circuits, *Proc. IRE*, May, 1947, pp. 519-525.

mobile receivers permeability tuning has technical and cost advantages which will probably cause it to remain in wide use.

Low-impedance loops can be permeability-tuned¹ by the use of (1) a loading coil, (2) an autotransformer, or (3) a two-winding transformer. With both transformer arrangements, the secondary is the tuning inductance which resonates with the fixed capacitor. Since the broadcast band embraces a frequency range of almost 3 to 1 and the tuning reactance must vary as the square of this, the inductance in a permeability tuner must have a wide range of variation. Typical quantities for a loading-coil design are as follows:

Max inductance of loading coil, μh	1,400
Min inductance of loading coil, μh	100
Effective permeability of core.....	12-15
Q of loading coil.....	70
Loop inductance, μh	7-20
Fixed tuning capacitor, μmf	100

38. Antenna Couplings for Sets Accommodating Open Antenna Only. The m-

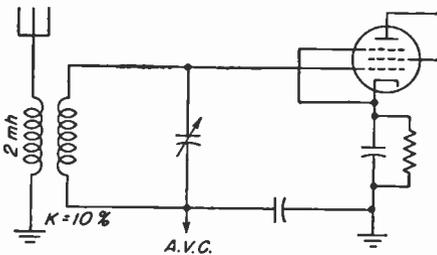


FIG. 5. Antenna transformer with high-inductance primary.

broadcast receiver designed for the weakest signals, such as in remote rural areas, must rely on a good open antenna. An effective height of 4 m, which is typical of outdoor antennas, is many times greater than obtainable with loop antennas. If few or no strong signals are expected, it is not worth while to provide a loop antenna. The design problem then is to provide good voltage gain and selectivity and also good alignment with other tuning circuits, all for any antenna capacitance within a range of 100 to 300 μmf .

From a chronological standpoint this problem was solved before loop antennas became common, and the result has been in wide use for a long time. The solution is shown in Fig. 5, where the 2-mh primary winding is sufficient to resonate the primary circuit below 550 kc for any value of antenna capacitance. Coupling of about 10 per cent furnishes enough mutual inductance to transfer adequate signal into the secondary circuit. In the absence of any unnecessary damping, it is necessary to avoid having the primary resonant in the

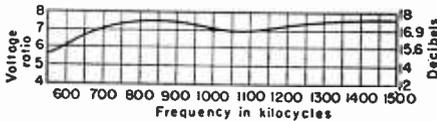


FIG. 6. Voltage gain obtained with antenna transformer having high-inductance primary.

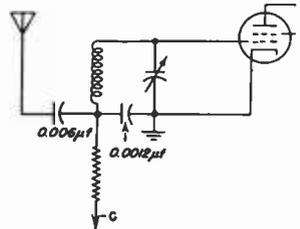


FIG. 7. Closely coupled antenna-input system with common low-side capacitor.

band, as the unicontrol alignment would be seriously affected. The relatively flat voltage-gain curve obtained with this type of transformer, Fig. 6, shows the ratio of voltage on the first grid to the voltage introduced in series with the standard dummy antenna.

The antenna-input system shown in Fig. 7 provides considerably greater coupling between the antenna and the first tuned circuit. This system is employed in automobile receivers where the signal intercepted by the antenna is usually quite small. By connecting a small inductance in series with the antenna so that a series-tuned cir-

¹ Vladimir, L. O., Permeability Tuning of Broadcast Receivers, *Electronics*, August, 1947, pp. 94-99.

cuit is formed which is resonant at approximately 2,000 kc, this system will provide a voltage gain which varies from 20 db (10 times) at 600 kc to 26 db (20 times) at 1,400 kc.

Another antenna input system which is used extensively in automobile receivers, particularly those designed for a specific car and antenna, is to connect the antenna to a tap at approximately 30 to 50 per cent on the coil in the first tuned circuit.

Permeability tuning for the antenna and r-f interstage couplings of automobile sets is now the general rule because of the economical construction, freedom from microphonism, and convenient adaptability to push-button tuning.

R-F AMPLIFIERS

The general underlying theory of r-f amplifiers is given in Chap. 9. The present section therefore covers some practical points regarding r-f amplifiers in radio receivers.

39. Inclusion or Omission of R-f Amplifier. The majority of home broadcast receivers now in use are five-tube, two-gang, a-c-d-c superheterodynes, as this design has been found the most popular combination of low price and acceptable performance. A set of this kind has no r-f stage, the five tubes being (1) the converter, (2) the i-f stage, (3) the diode detector and first audio stage in one envelope, (4) the output stage, and (5) the power rectifier. The sections of the two-gang variable capacitor tune the antenna coupling and the oscillator section of the converter.

Important advantages afforded by the use of an r-f amplifier with three-gang tuning include prevention of radiation of the oscillator output, reduction of the effect of the high tube noise of the converter, and a large improvement in the rejection of interference at the image and intermediate frequencies.

In special-purpose receivers, such as used in amateur, police, and point-to-point services, one or two tuned radio stages are generally provided. The high performance of the low-power types of the miniature and subminiature tubes makes them very attractive for consideration where the cost is justified. For example, the miniature 6AG5, 6AK5, 6AU6, 6BA6, and 12BA6 pentode types have transconductances in the range from 4,400 to 5,200 μmhos . The subminiature types, developed from the proximity-fuse work of the war, offer still smaller size with transconductances exceeding 3,000 μmhos .

Broadcast receivers include a tuned radio stage much more often in automobile sets than in home sets.

An untuned r-f stage used occasionally in medium-sized chassis for home receivers affords the advantages of suppressing oscillator radiation and giving some gain, at the cost of one tube and a very few inexpensive additional parts. Resistance coupling is generally used in such stages, giving a stage gain of 10 to 13 db (three to six times voltage gain) in the m-f band. If h-f international broadcasts are to be received, a high-transconductance tube must be used in the r-f stage to obtain appreciable gain, and it may be advisable to insert a series coil to resonate with the capacitances of the two tubes at a frequency near the upper limit of the range to be received. For example, a coupling to go up to 16 Mc and work between 10- μmf tube capacitances may include a coil to resonate at this frequency with the 5- μmf series equivalent of the two capacitances. This requires an inductance of 20 μh . A coil of medium quality is sufficient because high Q will give too sharp a peak at the higher frequencies of the desired tuning range.

A series trap for rejection of i.f. may be considered as an improvement for an untuned stage. This has the additional advantage of markedly improving the stability, *i.e.*, freedom from regeneration and tendency to oscillate, of the i-f amplifier of the receiver. Typical trap constants for 455 kc are in the neighborhood of 1 mh and 125 μmf , or 2 mh and 62 μmf . Factory adjustment and provision for field adjustment are needed. A high Q for this coil is desirable.

40. Inductive Tuning. Automobile broadcast receivers generally employ inductive tuning of the type in which a movable core of compressed powdered iron is inserted varying distances into the coil form to give the required inductance. Such cores are

called "slugs" In comparison with capacitive tuning in automobile receivers, this inductive type has been found more economical, more compact, and less subject to microphonics.¹ A typical interstage coupling of this kind has an almost constant Q with a value of about 80 over the band. The inductance range must be about 10 to 1 to cover the range 1,600 to 550 kc. A typical inductance range is 30 to 300 μh resonating with a fixed capacitor of about 250 μmf .

41. Unicontrol Tracking. The superheterodyne with an r-f stage normally requires the alignment of three tuned circuits of quite different characteristics. The r-f interstage coupling is taken as reference, in comparison with which the antenna coupling must accommodate a fixed loop antenna or an open antenna usually having any capacitance within a wide range. The oscillator circuit must, of course, operate at a frequency lying above or below the signal frequency by an interval equal to the i.f. Procedures for proportioning the oscillator circuit are given in Chap. 5 and in Secs. 58 to 60 of the present chapter. For push-button tuning refer to Sec. 61, *et seq.*

In the m-f broadcast band the usual tuning reactors for an r-f stage with capacitive tuning are a coil of about 200 μh and a variable capacitor with a maximum of about 420 μmf . With inductive tuning, typical values are a fixed capacitor of about 250 μmf and a moving-core coil with inductance range from 30 to 300 μh . The capacitors in inductive tuning are normally not actually fixed but are adjustable for alignment purposes.

In t-r-f receivers of former years, two or more similar interstage circuits were aligned for unicontrol. The method was to hold a close tolerance on the inductances of the coils and to provide trimming capacitors on the variable capacitor gang so that the minimum circuit capacitance of all circuits (with the gang entirely unmeshed) could be equated. With commercial capacitors this was usually sufficient, but sometimes the rates of change of the sections of the gang differed sufficiently, as they were progressively meshed, so that it might be necessary to bend the outer rotor plates of certain sections away from the adjacent stator plates. Normally the number of rotor plates exceeded by one the number of stator plates so that bending was possible at both ends of each capacitor section, and the outer rotor plates were often cut radially into sectors to facilitate this alignment adjustment if found necessary.

42. Cross Modulation in R-f Tubes. When a receiver is tuned to a strong signal and there is also present another strong signal of not too different frequency, cross modulation or cross talk may occur. During intervals when there is no modulation on the desired carrier, the modulation of the undesired carrier is heard, but if the desired carrier is absent, the undesired program disappears. The difficulty may occur in an r-f stage in either a superheterodyne or a t-r-f receiver. The cause is the operation of the tube with a total input so large that the linear part of its characteristic is exceeded. Under these circumstances, the tube acts as a modulator and produces a variation in the amplitude of the desired signal in accordance with the modulation of the undesired signal; *i.e.*, the tube gives to the desired signal a modulation from the undesired signal, whence the name "cross modulation." Since the interference becomes part of the modulation of the desired carrier, no subsequent operations can remedy the condition. No special frequency relation exists between the desired and undesired carrier frequencies. In this respect the phenomenon differs from external cross modulation (Sec. 30) and from spurious superheterodyne responses (Sec. 47).

To avoid this fault, the gradual-cutoff, or "variable- μ ," type of tube has come into use. Such a tube, if preceded by a single tuned circuit, is normally sufficient to prevent difficulty on this account.

In severe cases of cross modulation one undesired program may be modulated onto any one of several stations tuned in, so that any one of these stations selected is heard with the accompaniment of the undesired program.

The degree to which a given tube will produce this interference can be measured by noting the amount of increase of transconductance when the tube is measured at a low

¹ Polydoroff, W. J., Permeability Tuning. *Electronics*, November, 1945, pp. 155-157. Martowicz, C. T., Powdered Iron Cores. *Electronic Inds.*, June, 1945, pp. 108-110.

level of signal and then at a substantial level.¹ Another method, described in the same reference, is to measure the tube with a low value of signal and then superpose a substantial signal of considerably higher frequency, again noting the increase of transconductance. The greater the increase of transconductance, the more cross modulation the tube will produce.

43. Superheterodynes. Practically all receivers now being made are superheterodynes. This applies to the wide range from low-price broadcast receivers to excellent special-purpose designs.

The advantages of the superheterodyne are primarily the advantages of i-f stages in comparison with t-r-f stages. The much greater amplification and selectivity per stage realized at i.f. is (1) because no provisions have to be made in the i-f amplifier for variable tuning for the selection of stations, and (2) especially because the value of the i.f. can be chosen for the best performance. An additional advantage of the superheterodyne is the increase of stability afforded by the use of an additional frequency range; gains that would almost surely cause oscillation in a t-r-f receiver are easily obtained with the superheterodyne.

44. Analysis of Typical Superheterodyne Performance. The results normally obtained with a good grade of superheterodyne broadcast receiver having two i-f stages are shown in Figs. 8 and 9 in terms of total voltage gain and selectivity.

The performance of any one stage is shown by the ratios of the ordinates of the proper two curves at the various frequencies. For example, in Fig. 8 the voltage gain in the second i-f stage at 1,000 kc is total gain to the second detector divided by the total gain to the second i-f grid, or 960,000/16,000, which is 60 times or 36 db. In a similar way the selectivity of each stage at any frequency is obtainable from Fig. 9.

Typical over-all performances of broadcast receivers having various numbers of tubes, power supplies, and antenna provisions, are given in Sec. 23.

45. I.F. Higher Than Signal. In a few cases it is desirable to have the i.f. higher than the signal, i.e., to raise the incoming signal in frequency rather than lower it. Such a circuit reduces the wavelength of the received signal and therefore has been called an "infradyne."

¹ Herold, E. W., Simple Method for Checking R-f Distortion or Cross Modulation of Pentode Amplifier Tubes, *Electronics*, April, 1940, pp. 82-88.

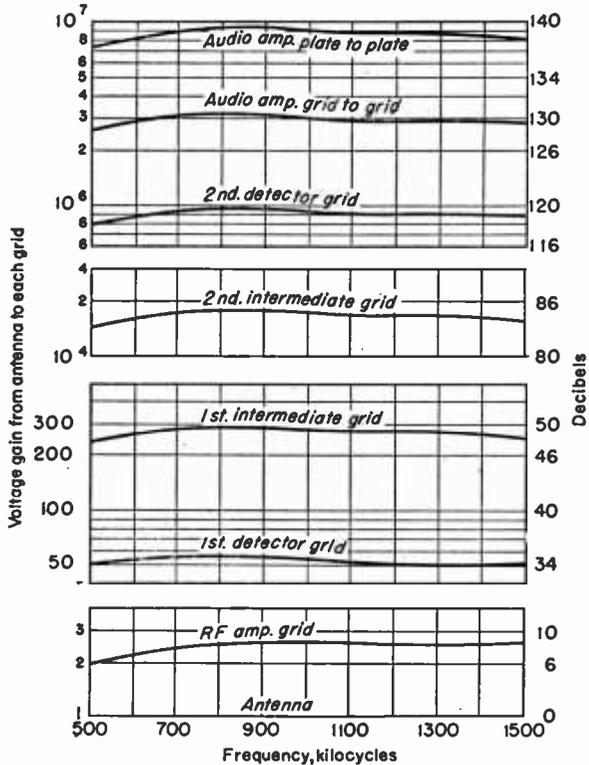


Fig. 8. Voltage gain in successive stages of superheterodyne receiver.

An instance is the 500-kc autoalarm set¹ of the Radiomarine Corporation of America where good image rejection is obtained by the use of a 1,100-kc i.f. The oscillator operates at 1,600 kc, and the image is at 2,700 kc. An autoalarm set maintains a watch for the international distress signal at ship and land stations when the operator is off duty or doing other work.

A number of broadcast receivers have been made in Europe, in which the use of an i.f. in the neighborhood of 2,000 kc allows the reception of both the l-f and m-f broadcast bands on one range of tuning capacitance.

The name "single-span"² has therefore been given to this type. With an oscillator range of 2,200 to 3,600 kc and an i.f. of 2,000 kc, all signal frequencies from 200 to 1,600 will be on the dial. A drawback is that the stations will be quite close together so that tuning may be difficult.

The use of an i.f. above the signal range gives excellent image rejection but is subject to spurious responses as follows:

Interference Due to Harmonics of the Received Signal. If the tuning range includes a signal frequency equal to one-half or one-third the i.f., such a signal may produce harmonics in the first detector which will be amplified by the i-f amplifier. I-f signals are thus produced without the use of the heterodyne oscillator. The frequency of the signals produced in this

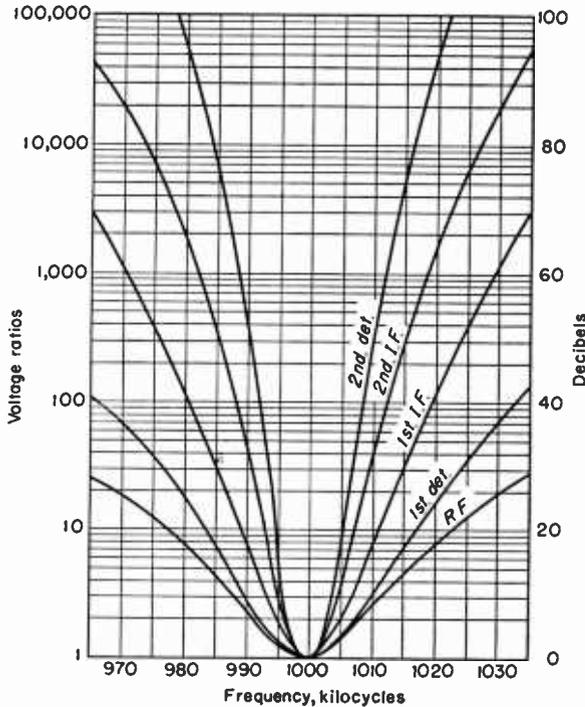


Fig. 9. Increasing selectivity in successive stages of superheterodyne receiver.

way does not vary as the receiver is tuned. The local oscillator also heterodynes the signal to the i.f., but the i.f. thus produced varies as the receiver is tuned. When the receiver is tuned through such a signal, a beat note, also called a "whistle," or "tweet," or "birdie," is produced by the two i-f signals. Selectivity ahead of the first detector will restrict the tuning range over which this interference is encountered but cannot eliminate it when the desired signal is the signal causing the interference; i.e., whistles will be heard when endeavoring to receive signals near subharmonics of the i-f.

Interference Due to Two Signals Whose Sum Frequency Equals the I.F. When two signals are impressed on the first detector and produce a sum frequency equal to the i.f., a beat note is produced as the receiver is tuned through a desired signal. Under this condition two i-f signals are produced, one of which remains fixed in frequency while the other varies as the receiver tuning is changed. Since the signals which can produce this interference may be on adjacent channels, the selectivity which must be used ahead of the first detector to avoid this interference entirely is equivalent to that normally used in the complete receiver.

¹ Byrnes, I. F., and H. B. Martin, Automatic Alarm, *RCA Rev.*, January, 1937, pp. 49ff.

² Colebrook, F. M., Frequency Transformation with Reference to Single Span, *Wireless World*, Feb. 15, 1935, p. 174.

46. Double Superheterodynes. For special purposes there is often a need for repeating the frequency change an additional time. This is especially the case in the v-h-f range. In a double superheterodyne there are two i.f.'s, the first of which is usually the higher and is chosen to give the required image rejection. The second i.f. is then chosen for best amplification and close-in selectivity considering the band width necessary to handle the signal. The first i.f. is sometimes adjustable as an additional means of station selection. Since a double superheterodyne has two frequency converters and a detector, it is sometimes called a "triple-detection" receiver.

The value of the first i.f. in double superhets usually lies between 1.5 and 50 Mc for services such as single and multiplex phone. (For radar, where a very wide band width is necessary, the single superheterodyne is generally used.) The second i.f. ranges ordinarily from 455 to 5,000 kc.

The disadvantages of the double superhet are (1) the added complexity and (2) the tendency to give numerous and prominent spurious responses. Considerable design and development work are therefore generally required to produce a model with acceptable performance.

An unusual but simple double superheterodyne was produced by Bendix,¹ in which the lower i.f. preceded the higher. The first i.f. was 9.5 Mc, and the second was obtained by doubling this to 19 Mc. Since a doubler, as in transmitter practice, requires no heterodyne oscillator, the frequency change was easily accomplished and met the desired need for increased stability by the use of another frequency.

One distinct type of double superheterodyne uses the same oscillator frequency for both conversions. As an instance, the model AN/TRC-1 series of f-m radiotelephone equipments² for a range of 70 to 100 Mc have a double-superhet receiver with a single oscillator whose frequency can be adjusted from 37.5 to 52.5 Mc. The first i.f. is variable in all single-oscillator double superheterodynes if a tuning range is provided, and in the present instance it may have any value from 32.5 to 47.5 Mc. The second i.f. is fixed at 5 Mc. Assume that a signal of 80 Mc is to be received, for which the oscillator is set at 42.5, the first i.f. is 37.5, and the second i.f. has its constant value of 5 Mc. As another example, for receiving a 100-Mc signal the oscillator is at 52.5 and the first i.f. is 47.5 Mc. If f is the signal frequency to be received and F is the fixed second i.f., then the oscillator frequency f_o and the first i.f. f_1 are as follows:

$$f_o = \frac{(f + F)}{2} \quad \text{and} \quad f_1 = \frac{(f - F)}{2}$$

These formulas give oscillator frequencies higher than the first i.f.'s; if oscillator frequencies lower than the first i.f.'s are desired, the formula for f_1 can be used to get the oscillator frequency, and the formula for f_o to get the i.f.

A multiple superheterodyne of special interest was used on Jan. 10, 1946, by Lt. Col. John H. DeWitt, Jr., and four assistants of the Signal Corps to receive the first radar echo from the moon.³ The achievement took place at the Evans Signal Laboratory, Belmar, N.J. To obtain the very weak echoes of 0.002 to 0.02 μ v per m at the antenna, corresponding to 0.01 to 0.1 μ volt delivered to the preamplifier, it was necessary to reduce the effect of noise by greatly restricting the band width and providing a receiver of very low noise figure. The equipment consisted of a much modified SCR-271 radar in which the original operating frequency of 111.5 Mc was retained. A band width of only 57 cycles was used, which was sufficient for the pulse of $\frac{1}{4}$ sec. duration sent out every 4 sec. The trip to and from the moon requires a total of 2.4 sec. To realize the necessary narrow band width, very precise control of frequency was required, and in addition the Doppler change of frequency on reflection at the moon had to be taken into account. This effect amounted to 300 cycles and changed in sign according to whether the observations were being made at moonrise or moonset. The most practicable way of meeting the requirements as to frequency stability and band width with the available equipment was to employ a quadruple superheterodyne, the i.f.'s being approximately 33 Mc, 6.7 Mc, 1.6 Mc, and 180 cycles. The 57-cycle band width was obtained with a pass band of approximately 150 to 210 cycles for the last i-f amplifier centering at 180 cycles. The echo from the moon was seen on a scope and also heard as a pulse of 180-cycle tone every 4 sec.

¹ Martin, David W., I-f Amplifier for High-gain F-m Receiver. *Tele-Tech.*, February, 1947, pp. 60-62.

² Marks, W. S., Jr., O. D. Perkins, and W. R. Clark, Radio-relay Communication Systems in the United States Army. *Proc. IRE*, August, 1945, pp. 502-522.

³ Mofenson, Jack, Radar Echoes from the Moon, *Electronics*, April, 1946, pp. 92-98.

The antenna had a gain of 23 db over an isotropic radiator. The preamplifier of the receiver was a special design with a noise figure of only 5 db. It had three stages of which the first two were grounded-grid 6J4 triodes, a miniature type with the high transconductance of 12,000 μ mhos. The noise power was doubled by the acceptance of the image in the frequency conversion from 1.6 Mc to 180 cycles, so that the noise figure of the entire receiver was 8 db.

47. Spurious Responses in Superheterodynes. The selectivity of a superhet in terms of selectances and of band widths at certain input ratios is not a complete indication of the actual selectivity under all conditions, as this type of receiver is susceptible to certain kinds of interference which are not encountered with a t-r-f receiver.¹ The susceptibility to these interferences is a result of frequency conversion. Considerable care may be necessary to distinguish interference of this general type from that due to cross modulation or external cross modulation.²

The discussion in this section is primarily for the case where the i.f. is lower than the received frequencies. For spurious responses when the i.f. is above the received frequency range, see Sec. 45.

48. Image Interference. The most important spurious response in a superheterodyne is usually the image. With the set tuned to a particular desired signal, the image lies on the other side of the oscillator, each being separated from the oscillator by the i.f. For example, a receiver tuned to 1,000 kc and having an i.f. of 455 kc is subject to image interference at 1,910 kc. Any receiver with this value of i.f., when tuned to any station, is subject to image interference at a frequency 910 kc higher than the desired station.

Since the function of the converter is to produce the difference between applied frequencies, it cannot distinguish between the signal and the image and produces i-f output from each. The only remedy for image interference is increased general r-f selectivity ahead of the converter, or special circuits³ designed to have a high attenuation at the image frequency, wherever it may be, as the tuning dial is adjusted to select any desired station. The simplest means for obtaining image rejection is to tap down the output lead on the tuning inductor. This can be done to give theoretically infinite attenuation at one frequency and substantial improvement over a band. The tap should be located at N_m turns from the low-potential end of the coil according to the formula

$$N_m = N_s (1 - f_s^2/f_m^2)$$

where N_s is the total number of turns, f_s is the signal frequency, and f_m is the image frequency. For a 1,000-kc signal and a 1,910-kc image the tap is at 75 per cent of the turns from the ground end of the coil.

49. Interference at the I.F. If there is interference at the i.f. and if an appreciable part of it reaches the converter, it will go through without change of frequency and appear as interference in the output of the receiver. In the United States, where 455 kc is the usual i.f., such interference is likely only from ship telegraph traffic of foreign vessels near our shores and of vessels operating in the Great Lakes. The FCC has been informed of the standardization on 455 kc and is expected to grant no new licenses to transmitters for operation at or near this frequency, although it may continue to renew old licenses.

The provision of traps in receivers for removal of i-f interference has an advantage in improving the stability of the i-f amplifier as well as removing interference. Such a trap is therefore included in an appreciable number of models now manufactured. It takes the form of a series-resonant circuit connected from the high side of the r-f circuit to ground.

¹ Morgan, Howard K., *Interfering Responses in Superheterodynes*, *Proc. IRE*, October, 1935, pp. 1164-1170.

² A useful comprehensive table, "Analysis of Radio Interference Phenomena," appeared in *RCA Radio Service News*, February, 1946, and was reprinted in *Radio News*, June, 1946, p. 54.

³ Wheeler, Harold A., *Image Suppression in Superheterodyne Receivers*, *Proc. IRE*, June, 1935, pp. 569-575.

50. Whistles or Tweets. A whistle or a "tweet" is a disturbance,¹ which consists of an audio note whose pitch varies in a sirenlike manner, passing from a high audio frequency down through "zero beat" and back up again as the dial of the set is rotated. All tweets result from the existence of two slightly different i.f.'s at the input of the second detector, the two frequencies having a difference which varies with change of the local-oscillator frequency so that the audio note goes through the characteristic progressive change of pitch. In some cases one of the i.f.'s is constant, and the other varies as the oscillator frequency of the set is changed. In general, both i.f.'s vary and this may be either in opposite directions or in the same direction at different rates. Whistles and images arise in the same manner, a whistle being essentially an image response in which the carrier phenomena, as distinct from effects due to modulation, are the chief noticeable characteristic. In whistles the effects of the side bands of the interference are negligible in comparison with the effects of the carrier.

Whistles may also occur when one is listening to a station near a harmonic of the i.f., without the presence of any interfering signal. With the usual 455-kc value, such whistles are likely to mar reception of stations on 910 kc and near 1,365 kc. If good reception of these frequencies is desired, the i.f. can be shifted slightly. A whistle can exist at such a point because two values of i.f. are produced when the set is slightly detuned. Assuming that the desired signal is 910 kc and that the oscillator is detuned to 1,366 kc instead of 1,365, these two i.f.s are (1) $1,366 - 910 = 456$, the i.f. representing the desired program, and (2) another due to demodulation between the second harmonic of the signal and the oscillator, $2 \times 910 - 1,366 = 1,820 - 1,366 = 454$ kc. A whistle of 2 kc is heard.

A whistle can be measured by adjusting it to approximately 400 cycles and observing under definite conditions the per cent of modulation of the desired signal produced by the whistle. Such measurements are made in receiver design laboratories, and the following are about the maximum values considered to be allowable:

Signal input at antenna terminals.....	0 db 1 volt	20 db 0.1 volt	40 db 0.01 volt	60 db 1,000 μ v	80 db 100 μ v	100 db 10 μ v
Whistle modulation, per cent.....	15	3	1	Not measurable		

If larger values are found at the larger signal inputs, the cause is known from experience to be overloading of the converter. As to the voltage gain ahead of the converter, low whistles on strong signals and low noise on weak signals impose conflicting requirements. A good compromise for the m-f broadcast band for a set with a tuned r-f stage is a voltage gain of 14 db (5 times) in the antenna coupling and a gain of about 20 db (10 times) in the r-f stage.

If a measurable whistle at 910 kc is found for the low inputs such as 80 db (100 μ v), the cause is a feedback of harmonics of the i.f. from the second detector into the r-f circuits. The remedy is to rearrange the wiring or provide shielding or by-passing to remove the feedback. The feedback may be through the a-v-c bias lead.

51. Sources of Whistles. In severe cases whistles occur in large numbers. Analysis indicates that there are numerous interfering frequencies which will satisfy a general whistle formula, as follows:

$$|mH - nS| = f_i$$

where m and n are small integers, H is the heterodyne oscillator frequency, S is the interfering signal causing the whistle, and f_i is the i.f. Values of S for a 455-kc i.f. and various ms and ns are plotted in Fig. 10.² Whistles are worse under the following conditions: (1) limited selectivity ahead of the converter, (2) small values of m and n involved, and (3) strong interfering signals.

Curve 13 is included in Fig. 10 for interference equal to the oscillator frequency.

¹ Strutt, M. J. O., Whistling Notes in Superheterodyne Receivers. *Wireless Eng.*, April, 1935, pp. 194-197.

² Barton, Don, Beat Frequency Interference Chart, *Electronics*, April, 1946, p. 162.

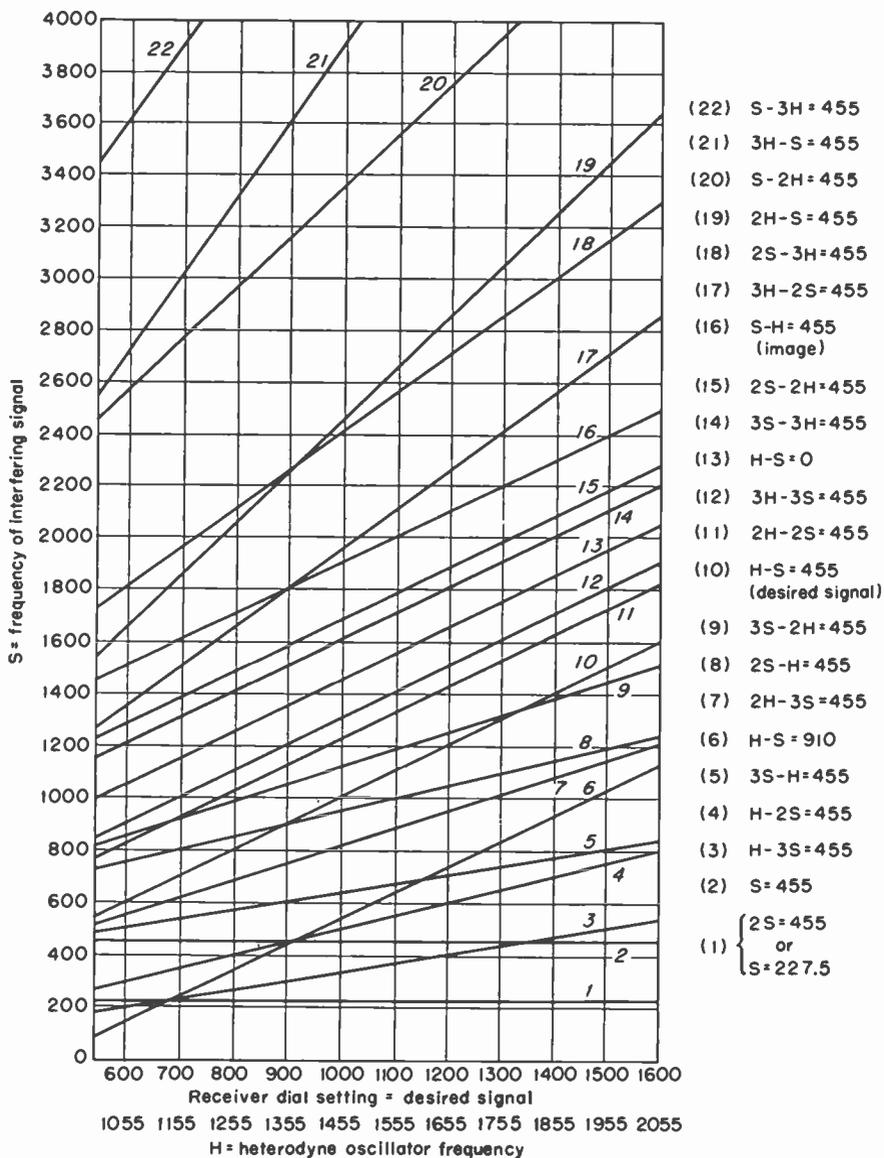


FIG. 10. Chart for identifying whistles heard with superheterodyne receivers having an intermediate frequency of 455 kc.

Obviously such interference amounts to another oscillator, and another i.f. whose frequencies do not change as the receiver dial is slightly rotated. The regular i.f. will change with dial rotation, so a whistle will be produced.

Another special curve in Fig. 10 is for interference located 455 kc below the signal to which the set is tuned. Only in a case of a strong interfering signal would this be important. This is curve 6.

Curves 2 and 1 show the susceptibility of the receiver to whistles resulting from interference at the i.f. and at half the i.f. The frequency of the local oscillator does not enter here.

On the plot of the figure it is not possible to show whistles due to two strong stations whose frequencies differ by 455 kc. If the r-f selectivity is inadequate, these will beat in the converter and thus produce a fixed-frequency signal in the i-f amplifier. The normal i-f signal, whose frequency depends on the dial setting of the receiver, will therefore complete the requirements for a whistle. This will be heard on any desired station as long as the two undesired stations reach the converter in sufficient strength.

52. Tests of Spurious Responses. The 1938 IRE "Standards on Radio Receivers" include procedures for testing the amount of image and i-f interference and of whistles. These are briefly summarized in the following paragraphs.

The sensitivity of a receiver to image interference is measured in the same way as normal sensitivity except that the signal generator is adjusted to produce the image frequency. The observed value is the image sensitivity in decibels below 1 volt or in microvolts. Observations are made with the receiver tuned successively to one or more frequencies in the various bands that are provided. The image rejection is a comparison with the sensitivity of the receiver for the frequency to which it is tuned and is expressed in decibels or as a voltage ratio. As an example, a set with a normal sensitivity of 100 db and an image sensitivity of 60 db has an image rejection of 40 db.

I-f sensitivity is measured with the signal generator adjusted to deliver the i.f. and the receiver tuned successively to the various frequencies of interest. The i-f rejection is obtained in the same way as for image rejection.

The whistle-modulation test consists of a survey of the entire frequency band or bands of the receiver, using an unmodulated signal from the signal generator and keeping the receiver constantly in tune with the generator. At each whistle located a measurement of its per cent modulation is made by adjusting it to about 400 cycles and using the formula,

$$m = \frac{30E_w}{E_s}$$

where m = desired modulation percentage

E_w = output whistle voltage

E_s = signal output for 30 per cent modulation at 400 cycles

As an example, if E_w should equal E_s , the value of m would be 30 per cent, which would be reasonable as indicating equality to the 30 per cent signal modulation.

The test is repeated at representative signal levels. The results give a clear picture of the minimum whistle interference which characterizes the set, since in this test there is no interference to aggravate the situation. In this test, whistles will be found at harmonics of the i.f. and possibly at other points.

A two-signal test is also provided in the IRE Standards which simulates the practical case where the receiver is tuned to a given frequency but subject to interference over a wide frequency range. One of the signal generators supplies the input to which the set is tuned, and the other generator is used to explore for whistles. As each one is located, a measurement is made of the input of the "interfering" signal generator necessary for the whistle, when tuned to 400 cycles, to be 30 db below standard output. The results are plotted as a spectrum with frequencies as abscissas. The ordinates are decibels below 1 volt plotted linearly or microvolts plotted logarithmically.

53. Frequency Converters. The frequency-changing function is performed by a modulator, or mixer. This receives the r-f signal and also the output of a local oscil-

lator, and delivers an i-f signal. The general theory of modulators is given in Chap. 10. The modulator, or mixer, and the oscillator are generally called a "converter." The mixer and oscillator of a converter may be in either the same or separate tubes. Extensive studies have been made of various converters.¹ One classification of converters is according to whether the signal and heterodyne inputs are applied at the same electrode or at separate electrodes and, in the latter case, according to whether the electron stream is first influenced by the heterodyne or by the signal voltage.

54. Converter Performance in M-f Broadcast Band. A survey of recent receivers, mostly a-c-d-c table models, shows that the most widely used converter tubes are the octal 14Q7, which is a heptode with a conversion conductance of 525 to 550 μmhos , and the octal 12SA7, which is also a heptode but with a conversion conductance of 425 to 450 μmhos .

In another survey of more than 50 sets tested in the Hazeltine Laboratories during 1946, the average conversion gain for several classes of receivers lay in the range from 28 to 34 db. The d-c voltage drop in the oscillator grid leak, which is equal to the peak r-f oscillator voltage, was 9 to 10.5 volts. For design purposes, measurements were also made of the gain of the converter stage with i-f instead of r-f input applied on the signal-input grid; these indicated a gain of about 2 db more than the conversion gain and gave a band width at 6 db of 10 to 13 kc and at 20 db of 25 to 34 kc. These band widths show the selectivity of the coupling system between the converter and the first i-f stage.

55. Converter Performance on 100-Mc F-m Band.² Tests on ten 100-Mc f-m receivers designed during 1946 gave an average of 17.5 db conversion gain with an oscillator voltage about half that used in the 1-Mc broadcast band. Measurements of the gain of the converter stage made at 10.7-Mc i.f. gave an average value of 21 db, with a band width at 6 db of 240 kc and at 20 db of 610 kc.

In sets now made the most common converter tube is the 6SB7Y having a conversion conductance of 950 μmhos and a tube noise which is equivalent to only 5 μv on the grid.³ This is a considerable improvement over the 6SA7 and the 12SA7. The 6SB7Y is designed to operate with a much lower oscillator voltage than previous types, *i.e.*, only 0.5 to 1 volt, representing a current of 120 μa through a grid resistor of 6,800 ohms.

56. Germanium Crystal Converter for V.H.F. At frequencies up to 200 Mc germanium crystals⁴ offer the advantages of compactness, absence of heater power requirements, and good electrical characteristics. The 1N34 and 1N38 crystals have the following ratings:

	1N34	1N38
Frequency range, Mc.....	0-100	0-200
Peak inverse voltage.....	50	100
Max avg current, ma.....	22.5	22.5
Peak current, ma.....	60	150
Max transient or surge current, ma.....	200	500
Max back conduction at 50 volts, ma.....	2	$\frac{1}{4}$
Shunt capacitance, μmf	3	1

The 1N38 is stated to have a life of at least 5,000 hr at the rated maximum average current of 22.5 ma.

¹ Herold, E. W., The Operation of Frequency Converters and Mixers for Superheterodyne Reception, *Proc. IRE*, February, 1942, pp. 84-103. Cowan, F. A., Modulation in Communication, *Elec. Eng.*, September, 1947, pp. 919-924.

² Herold, E. W., and L. Malter, Some Aspects of Radio Reception at Ultra-high-frequency, *Proc. IRE*, October, 1943, pp. 587-582.

³ Operation of the RCA 6SB7-Y Converter, *Application Note* 120, RCA Tube Department, Harrison, N.J., June 18, 1947. See also Use of 6BA6 and 6BE6 Miniature Tubes in FM Receivers, *Application Note* 121 of the same date.

⁴ Cornelius, E. C., Germanium Crystal Diodes, *Electronics*, February, 1946, pp. 118-123. Silicon and Germanium Rectifiers, *Electronics*, Buyers' Guide Issue, June, 1947, pp. 146-147.

57. Silicon Crystal Converter for U.H.F. and S.H.F. In the range from 1,000 to 25,000 Mc, the fixed silicon crystal with tungsten point is generally used as converter because of its low noise and moderate conversion loss.¹ Below 2,500 Mc tube converters can be employed with the advantage of greater ruggedness but the drawback of poorer electrical performance.

Crystals must be protected from electrical overload, such as due to poor duplexer operation in radars. The burnout-test energy in a "spike" of admitted transmitter output is only 2 ergs for 3,000-Mc crystals, 0.5 or 1 erg for 9,000- and 10,000-Mc types, and 0.1 erg for the 25,000-Mc type 1N26. For 1,000-Mc service, the Western Electric 1N25, shown in Fig. 11, is available with the high value of 30 watts burnout pulse power.

As typical performance for the frequency range from 1,000 to 30,000 Mc, a conversion loss of 6.5 to 8.5 db can be stated, and a noise figure of only 1 to 3 db. As a

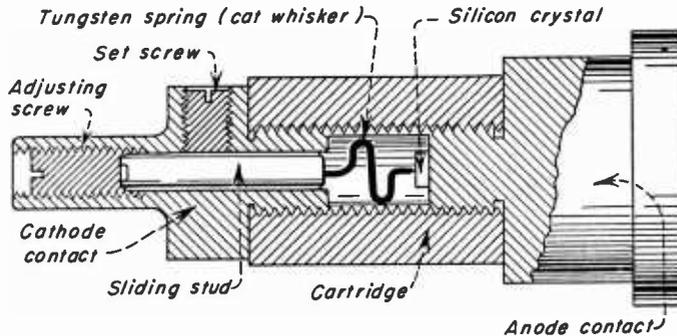


FIG. 11. Construction of 1N25 crystal designed for 1,000-Mc service. Crystal has high conductivity when anode is positive and cathode negative.

source of the output i-f signal, these crystals have an internal resistance of 150 to 800 ohms.

At lower frequencies crystals produce more noise for a given band width.² This is true below 1 Mc and especially in the a-f range.

58. Single-dial Tuning.³ A superheterodyne receiver involves two problems to achieve single-dial tuning. One, also present with the t-r-f receiver, is the tracking of the antenna coupling circuit to the interstage circuit in such a way as to accommodate the necessary range of antenna capacitances. The use of a high-inductance primary in the antenna transformer for this purpose is described in Sec. 38.

To secure satisfactory alignment, fixed tuning elements such as coils are usually made with very close tolerances. For this reason coils are sometimes wound in two sections, as shown in Fig. 12. With this construction one or more turns in section *a* can be transferred to section *b* so as to increase slightly the total inductance of the coil. It is evident that this will increase the inductance because in the limit, when all of *a* is transferred to *b*, the coupling between the turns formerly constituting *a* and the remainder of the coil is increased which increases the total coil inductance. Conversely, if the inductance of the coil is too high, it can be reduced slightly by transfer-

¹ Cornelius, E. C., Silicon Crystals for U-h-f Detection Circuits, *Electronic Inds.*, November, 1945, pp. 74-77. Miller, S. W., Considerations in the Design of Centimeter-wave Radar Receivers, *Proc. IRE*, April, 1947, pp. 340-351. Scaff, J. H., and R. S. Ohl, Development of Silicon Crystal Rectifiers for Microwave Radar Receivers, *Bell System Tech. J.*, January, 1947, pp. 1-30. Edwards, C. F., Microwave Converters, *Proc. IRE*, November, 1947, pp. 1181-1191. Pound, R. V., "Microwave Mixers," McGraw-Hill, 1948.

² Miller, P. H., Jr., Noise Spectrum of Crystal Rectifiers, *Proc. IRE*, March, 1947, pp. 252-256. Roberts, Shepard, Some Considerations Governing Noise Measurements on Crystal Mixers, *Proc. IRE*, March, 1947, pp. 257-265.

³ A table showing various general alignment procedures is given by Irving Dlugatch, Alignment Methods, *Radio Maintenance*, December, 1947, pp. 14-16.

ring turns from b to a . After this factory adjustment a suitable lacquer is applied to hold the turns in place and protect them from humidity. Instead of such close adjustment it is sometimes practical to sort coils in the factory according to their inductance and then use similar ones together; *i.e.*, if the tuning coils of all circuits are slightly high but in corresponding amounts, the single-dial tuning will not be disturbed beyond a slight permissible error in the frequency as indicated on the dial.

The superheterodyne, in addition to antenna tracking, introduces the problem of oscillator tracking, with the requirement that the oscillator circuits differ accurately from the signal circuits by just the amount of the i.f. The oscillator almost always is set at a frequency higher than the signal because this gives a smaller value for the ratio of maximum to minimum oscillator frequency. The smaller value of this ratio is desirable because it permits a much smaller variation of the tuning reactor in the oscillator circuit, which results in a cost saving.

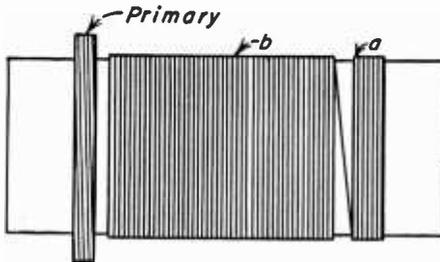


FIG. 12. Method of varying inductance slightly in factory for tracking purposes.

The oscillator circuit may consist of a 200- μ h coil and a tuning capacitor with a maximum of 420 μ mf. The oscillator section consists of an 80- μ h coil and a cut-plate section of the tuning gang with a maximum capacitance of 160 μ mf. The reduction of oscillator tuning capacitance saves chassis space owing to the shorter length of the capacitor gang and the smaller space required for the emerging rotor of smaller radius. A capacitor gang with a cut-plate section is generally avoided in receivers having more than one tuning band but is frequently found in sets receiving only the standard broadcast band.

Capacitor gangs with identical sections can be used by employing a suitable value of inductance in the oscillator and by employing series and shunt capacitors. A method of computing the proper inductance and proper values for the associated capacitors is given in Chap. 5. This gives a "three-point tracking"; *i.e.*, the alignment is perfect at three points in the band. The choice of the inductance, which is a design and factory matter, determines the frequency near the center of the band at which the tracking is perfect. The series capacitor furnishes perfect tracking at a point near the l-f end, and the parallel capacitor near the h-f end. These two capacitors are normally adjustable in the field whenever necessary to restore satisfactory alignment.

A simplified procedure for determining the oscillator inductance and the series and parallel capacitors has been described¹ together with an illustration for coverage of the h-f range from 7 to 15 Mc. Lumping all the minimum circuit capacitance together into a total of 20 μ mf and using a capacitor with a maximum of 83 μ mf with a signal tuning inductance of 5.6 μ h, the oscillator inductance was found to be 5.2 μ h, the series capacitor 1,370 μ mf, and the parallel capacitor 0.3 μ mf.

60. Tracking with Permeability Tuning. In automobile receivers, permeability tuning has practically displaced capacitor tuning. This brings up various questions with regard to the tracking of the oscillator.²

¹ Marshall, Joseph, Superhet Tracking Formulas, *Electronics*, October, 1946, pp. 202, *et seq.*

² Fairman, Harry E. Oscillator Tracking Methods in Permeability Tuning, *Tele-Tech.*, March, 1947, pp. 48-51.

There is a limited use of designs in which the variable inductor for the oscillator circuit is identical to the inductors in the signal circuits. In this case, a small series inductor followed by a shunt inductor and by the fixed tuning capacitor can be used to obtain three-point tracking. In a typical example,¹ the tuning inductors had a range from 120 μ h at 1,600 kc to 1,000 μ h at 550 kc, the signal tuning capacitance being 80 μ mf. In the oscillator circuit, the identical inductor was followed by a 26- μ h series coil, a 1,127- μ h parallel coil, and a 46- μ mf fixed tuning capacitor. The i.f. in this design was 455 kc.

Instead of the series and parallel inductors, it is often possible to obtain the effect of the series coil without actually having it present by the expedient of a thicker coil form, and to obtain the effect of the parallel inductor by the use of a tapered winding. This amounts, of course, to employing a different inductance coil in the oscillator circuit. Formulas applicable to this case have been given² in the literature.

PUSH-BUTTON TUNING

The convenience of bringing in favorite stations by merely pressing a marked button has made push-button tuning one of the standard features of broadcast receivers dur-

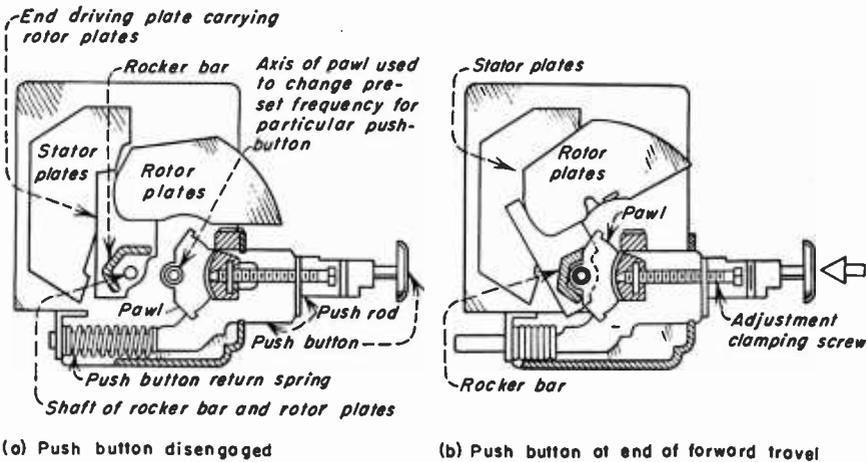


FIG. 13. Rocker-bar type of mechanical push-button tuner.

ing the last decade. Numerous methods of obtaining this result³ have been devised. Most automobile receivers now include push-button tuning, and many home receivers are also equipped.

61. Mechanical Rotation of Tuning Gang. One widely used method is to provide an arrangement of mechanical elements by which the user can push a particular button with sufficient force and through a sufficient distance to rotate the tuning capacitor to the required position. A stop establishing a limit of travel is provided to serve as an indication to the user that the capacitor is at the required position.

A widely used mechanical method of preset tuning has a semicircular pawl or disk which is advanced by the push-button, as shown in Fig. 13, the diameter edge of the pawl moving ahead in the direction of the button travel. This pawl encounters a

¹ Simon, A. W., Tracking Permeability Tuned Circuits, *Electronics*, September, 1946, p. 138. See also note by same author in *Electronics*, November, 1947, pp. 142, 156.

² DeCola, Rinaldo, A Superheterodyne Tracking Solution, *Electronics*, February, 1942, pp. 29-30, 91-92.

³ "MYE Technical Manual," P. R. Mallory and Co., Indianapolis, 1942. See Sec. 6, pp. 137-198. Williams, C. R. G., Tuning Devices for Broadcast Radio Receivers, *J. IEE*, Part III, November, 1946, pp. 405-428.

double-arm rocker bar which controls the position of the tuning gang. One end of the pawl encounters the arm on one side of the rocker bar and pushes it ahead until the other side of the rocker bar strikes the other end of the pawl, thus determining the end of button travel and the final position of the rotor of the tuning gang. In this type of mechanism, the preset station for each button can be changed by (1) releasing the adjustment clamping screw, which leaves the pawl free to rotate about its axis; (2) pressing the particular push button and holding it in; (3) tuning in the desired new station with the manual tuning control of the set; (4) tightening the adjustment clamping screw; and (5) releasing the push button. Capacitors with this preset mechanism are designed to go through their entire capacitance range in 90 deg of rotation, and good ratios of maximum to minimum capacitance are obtainable. If a 180-deg capacitor is to be used, a 2:1 stepup gearing is provided.

Another practical method of mechanical push-button tuning has heart-shaped cams on the shaft of the variable capacitor, one such cam for each preset station. Pressure on the push button brings a small roller against the cam and thus rotates the capacitor shaft until the roller ends its travel in the vee of the heart; here the roller is nearest the shaft and can go no farther, and the shaft has been rotated to the desired position. A clamping arrangement allows the cams to be loosened for resetting the stations assigned to one or more of the push buttons.

62. Mechanical Control of Slug Position. Push-button control of permeability tuners can be obtained by mounting the ends of the slugs on a common bar which then moves them into their respective coils. Another arrangement is to use cords to connect the slugs to the tuning dial in such a way that rotation of the dial pulls on one slug, the back end of which is tied to the next slug, the back end of which is tied in turn to the other side of the tuning dial.

63. Multibutton Self-disengaging Switch. Other types of push-button tuning are electrical in character so that the only function of each push button is to perform certain electrical connecting operations. One requirement on a gang switch of this kind is some type of latching mechanism to hold any button pressed and yet so arranged that pressing any button disengages whatever button was previously pressed. Moderate cost is, of course, necessary for these switches to be used in broadcast receivers, and in some types of circuits low capacitance is also required. The self-disengaging feature is obtainable with a latch bar by which the first effect of pushing any button is to disengage the previous button, after which further travel of the button makes the desired connections. The last button operated ordinarily stays more or less depressed to show the user what station is tuned in.

64. Single-button Switch. An extensive use has been made of a single button as the entire preset tuning control available to the user. This has been found acceptable especially in automobile receivers. Pressing the button causes the set to change from one preset station to the next preset station in a cyclical manner. The receiver in this case includes a solenoid-operated rotary switch which moves around one position for each operation of the single control button. An indicator, such as a numbered or lettered dial behind a window, indicates at all times what preset station is being received.

65. Station Choice by Preset Reactors. The most common electrical type of push-button tuning employs preset reactors for the choice of stations. Ordinarily one such reactor is provided for the tuning of the antenna coupling and one for the tuning of the oscillator. In sets having a tuned interstage coupling for continuous tuning, the r-f stage is usually switched out of the circuit for push-button tuning.

The best combination of stability and low cost appears to be the use of preset capacitors for antenna tuning and preset inductors for oscillator tuning. The function of the switch is to select the necessary pair of reactors to tune in the desired station.

66. Motor-operated Tuners. A limited number of sets with motor-operated tuners have been made in which an electrical motor is provided for rotating the tuning capacitor to the required position. The determination of this position, as shown in Fig. 14, is made by providing a disk, which is essentially a two-bar commutator, and several

narrow contact brushes, one for each preset station. A contact resting on one-half the disk will cause the motor to turn in one direction, while if it rests on the other half the motor will turn in the opposite direction. With this arrangement, whenever any button is pushed, the motor starts turning in the proper direction to rotate the capacitor shaft directly toward the desired point. As the capacitor rotates, the commutator moves until the insulation between the two halves of the disk is under the contact point, at which time the desired station is tuned in and the motor stops. The setting up of the chosen stations requires only that each contact be positioned around the disk at the necessary angle.

In practice, the contacts are usually arranged in two concentric guide channels. This permits setting two contacts to stations which may be close together in frequency. For setting the contacts, the desired station is tuned manually and the appropriate contact slid around its guide slot until it is on the insulating strip. A fairly high tuning speed can be provided since, if the momentum should cause the tuning to overshoot, the motor will automatically reverse and return to the correct point, where it will stop.

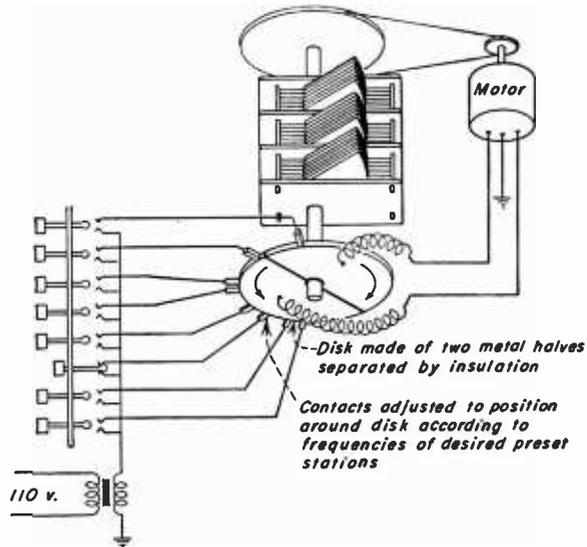


FIG. 14. Motor-operated preset tuner with control disk to start motor in the proper direction to go toward any desired preset station.

67. **Frequency Stability.** Stability of the various circuits, which is desirable in any receiver, is important for push-button sets. This matter has been the subject of much study and development, and the art is now in a satisfactory state in the sense that almost any desired precision can be obtained at moderate expense. In addition to components less subject to temperature drift,¹ compensating elements have been developed. Within the receiver the problem has been attacked by the use of a wider top i-f selectivity curve and in some elaborate receivers by the provision of a.f.c.

68. **Receiver Silencing.** Tuners which move the capacitor or inductor gang from one position to another should be provided with a contact for disabling the receiver during the time the change of tuning is in process. Otherwise a disagreeable series of sounds from the various stations passed through in the tuning will be heard. This does not apply to tuners that operate by making a quick change of preset reactors. A ground on the grid of an audio stage can be used to quiet the receiver where necessary during the change of tuning, the arrangement being such that silencing is accomplished in an automatic and economical manner.

69. **Remote Control.** Some push-button tuning systems lend themselves to convenient employment for remote control, if choice among preset stations, without continuous tuning, is sufficient. The single-button tuner is an outstanding case of this kind since only two wires are required for the tuning. Motor-operated types can also

¹ Batcher, Ralph, Thermal Stability in Receiver Oscillators, *Electronic Inds.*, April, 1945, pp. 96, et seq.; May, 1945, pp. 93, et seq.

be used for remote control but require a multiconductor cable. Further treatment of remote control is given in Sec. 105.

70. Preset Tuning at Higher Frequencies. Various preset tuners have been devised for aviation, maritime, and military equipment operating at frequencies above the standard broadcast band. In Western Electric equipment, a telephone dial is used to tune quickly to one of as many as 8 or 10 preset frequencies. Another tuner,¹ called the "Autotune," has been used extensively in military equipment. This is a motor-operated type which has for each preset frequency a circular disk with a gap into which a detent pawl enters, thus stopping the rotation of the shaft at the required point. It makes a change from one preset tuning to another in 5 sec or less.

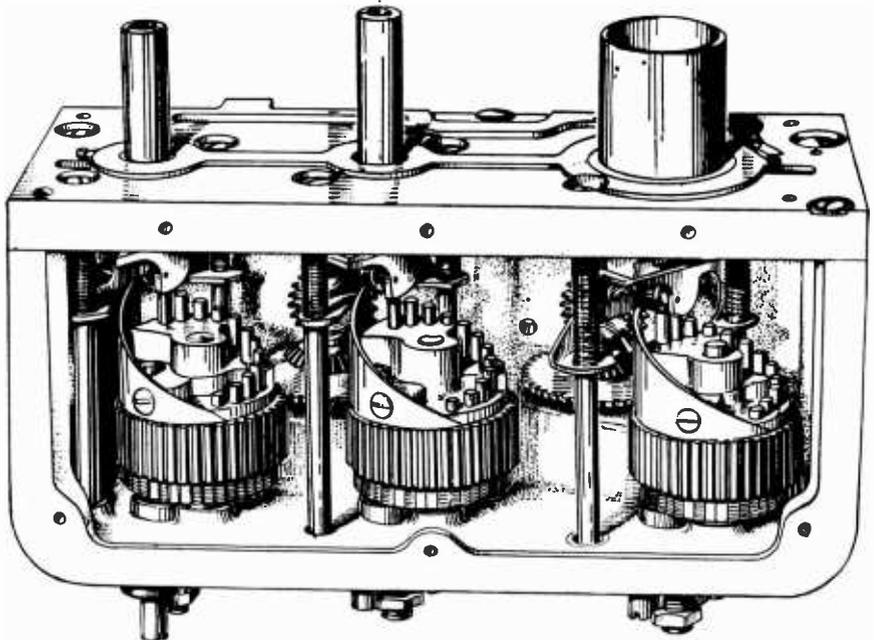


FIG. 15. Preset tuning mechanism for 1,000-Mc distance-measuring equipment.

Hazeltine engineers have designed a preset tuner for frequencies in the neighborhood of 1,000 Mc, which is carried in planes as a navigational aid for obtaining the distance from beacons located on the ground. In this application one knob affords the choice of any one of 13 transmitting frequencies by means of a detent tuning mechanism shown as the right-hand one of three such assemblies in Fig. 15. The upper ends of the detent pins have the appearance of a spiral staircase. The assembly rotates in only one direction, a plunger descending from one pin to the next until the desired one is reached. Whenever it is necessary to change to a higher pin, the one direction of rotation is continued, the plunger riding up to the top on the smooth curved surface shown and then coming down to the desired point.

The center and left-hand assemblies in Fig. 15 operate similarly, one controlling the oscillator and the other the preselector. A single shaft is provided to control both these assemblies, so that complete preset receiver tuning is afforded. In accordance with the usual requirement for the remote control of airplane radio equipment, provisions are made for remote choice of the desired transmitting and receiving frequencies.

¹ May, R. W., and N. H. Hale, Automatic Positioning Control Mechanisms, *Electronic Inds.*, January, 1946, pp. 58, et seq.

I-F AMPLIFIERS

Numerous factors must be considered¹ in the design of the i-f amplifier, especially equipment for special services. These factors include the choice of the intermediate frequency, number of stages, gain and selectivity required, whether variable selectivity is to be provided, whether stagger tuning is necessary, etc.

71. Choice of I.F. The chief considerations are (1) better image rejection is obtained with higher values and (2) better gain and selectivity in the i-f amplifier are obtained with low values. With broadcast receivers, the most widely used i.f. is 455 kc, which has been standardized by the RMA. A limited use is made of 262 kc in automobile receivers. For the f-m broadcast band (88 to 108 Mc) the i.f. of 10.7 Mc is generally used. In television receivers the value of the picture carrier in the intermediate stages is about 26 Mc, and the sound carrier is about 21.5 Mc. Radar receivers, with carrier frequencies in the u-h-f and s-h-f ranges, usually employ i-f amplifiers operating at 30 or at 60 Mc. A limited use has been made of i.f.s above 100 Mc.

One of the considerations bearing on the choice of i.f., especially for fixed-frequency operation such as in the point-to-point services, is to avoid values that might cause spurious responses due to high-order frequency conversions in the mixer tube.²

72. Typical I-f Amplifier Performance. A single i-f stage is usually sufficient in broadcast receivers, especially if only the standard broadcast band is to be accommodated. Such a stage is usually provided with input and output tuned transformers, each having tuned primary and secondary windings, the coupling being about 1 per cent and the Q about 90, so that the coupling is slightly under optimum, with the advantage that manufacturing tolerances do not result in any of the product having a double-peaked characteristic. The i-f gain of such a stage is ordinarily around 40 db, the band widths about 16 kc at 6 db and 44 at 20 db. The inductance of each winding is generally about 1.0 or 1.1 mh, requiring a capacitance of about 120 μ mf.

Sets accommodating the f-m broadcast band usually have 10.7-Mc i.f. and two intermediate stages. The first 10.7-Mc i-f stage usually has a gain of about 29 db and band widths of 240 kc at 6 db and 600 kc at 20 db. The second stage has a gain of about 26 db with greater band widths, representative values being 460 kc at 6 db and 2,100 kc at 20 db. Composite i-f transformers are widely used, in which one metal can houses both the 455-kc and the 10.7-Mc elements.

I-f transformers for 455-kc operation have recently appeared with greatly reduced physical dimensions—about $\frac{3}{4}$ by $\frac{3}{4}$ by 2 in.—and these with smaller tubes and other components tend toward a compact chassis.

73. Capacitive and Inductive Tuning. Capacitive tuning of 455-kc i-f transformers is more common than inductive tuning. The usual two-winding transformer is provided with two padders, one for each winding, which are screw-driver adjusted at the factory and can be changed later, if necessary. Greater stability as a function of temperature and age is obtained with inductive tuning in which fixed high-grade capacitors are used, and accurate tuning is obtained by adjusting the position of a powdered-iron slug in each winding. The presence of the iron core increases the Q of the coil so that the greater cost is partly offset by the fact that solid wire rather than stranded can generally be used.

74. Choice of Inductance. To obtain maximum gain in an i-f stage, the L/C ratio should be as large as possible with stability. Too high a value, obtained with small values of C , also makes the variations of the input capacitance of the following tube produce excessive misalignment of the tuning circuit. For this reason values of C are seldom below 30 or 40 μ mf, corresponding at 455 kc to 4 and 3 mh. Battery receivers, which must have high gain with limited plate voltage, use inductances between 2 and 3 mh.

¹ Spaulding, F. E., Jr., Design of Superheterodyne Intermediate-frequency Circuits, *RCA Rev.*, April, 1940, pp. 485-495.

² Charts for assisting in this evaluation have been prepared by R. S. Badessa, Mixer Frequency Charts, *Electronics*, August, 1946, p. 138.

With sets having two i-f stages, and especially with a-m-f-m receivers where high-gain tubes and three i-f stages may be necessary in f-m reception and where it is desired to avoid the switching that would be required to reduce the number of i-f stages in a-m reception, lower gain on the 455-kc i.f. is needed to avoid instability, and inductance values as low as 200 μ h are used.

Unequal primary and secondary inductances, but with equal primary and secondary Qs, are recommended for 10.7-Mc operation.¹ In this way only the tube and other distributed capacitance need be used to tune the primary, while 50 μ mf or more can be used for the secondary. Under these circumstances the primary circuit will give high gain and the secondary circuit will accommodate the change of as much as 2 μ mf in the input capacitance of the following tube over the range of a-v-c action. A coupling slightly under optimum is recommended.

As i.f.s of 10 Mc and higher are considered, the gain which can be obtained is limited by one of three factors.² One of these is band width, the difficulty being that if the gain is increased beyond a certain point, the band width will be narrowed; *i.e.*, for a given band width, other conditions being the same, there is a limit to the gain. The second limiting factor is stability, which depends on the grid-plate capacitance. If the gain exceeds a certain amount, sufficient voltage will be fed back to produce regeneration or oscillation. The third factor which may limit the gain is the input loading of the following stage. The losses due to transit time, cathode-lead inductance, and other causes are equivalent to a relatively low resistance shunted from grid to ground across the input of the following stage. This limits the gain that can be obtained.

75. Special Broadcast Receiver Designs. I-f transformers of various special kinds have been proposed or used in limited numbers.

Single-tuned transformers, with very closely coupled windings one of which is tuned, are less expensive than double-tuned designs and can be used where a limited amount of selectivity is sufficient. A triple-tuned design, having three windings and three capacitors, gives very good performance with a relatively wide flattop and sharp skirt selectivity. This type has been used to a limited extent with push-button receivers, the wide top accommodating some drift.

An interesting i-f coupling of a resonant mechanical type³ has been described, in which the mechanical equivalent of a band-pass filter centered at 455 kc is obtained by the use of stainless-steel plates coupled together by two parallel steel wires spot-welded to the plates. A typical design has a pass band of 8.5 kc, but affords an attenuation of 60 db at 10 kc away, *i.e.*, with a band width of 20 kc. Coupling into and out of the mechanical filter is by means of nickel magnetostrictive elements.

Another special class of i-f transformers is made for variable selectivity, usually having an additional winding which can be switched in or out to vary the coupling and thereby the band width. Further data on variable selectivity are given in Sec. 151.

For very sharp selectivity, such as for telegraph reception, a quartz-crystal filter is often used in the i-f amplifier. A description of this feature is given in Sec. 176.

The use of band-pass filters, including M-derived sections,⁴ affords sharp skirt selectivity for receivers in point-to-point practice. The over-all band width of one such amplifier measured 12½ kc at 6 db and only 19 kc at 90 db.

76. Proportioning of A-v-c Bias. In intermediate amplifiers of more than one stage, it may be advisable to reduce the amount of a-v-c bias applied to the last stage. This stage has the function of delivering the full amount of signal desired at the detector and should, therefore, not be biased back farther than this function permits. Otherwise distortion will be introduced, especially for high values of modulation.

¹ Aram, N. W., L. M. Hershey, and Marvin Hobbs, F-m Reception Problems and Their Solution, *Electronics*, September, 1947, pp. 108-113; see p. 112.

² Notes on Factors Limiting Gain in I.F. Amplifier Tubes, *Engineering News Letter* 88, Sylvania Electric Products, Inc., Mar. 15, 1947.

³ Adler, Robert, A Compact Electromechanical Filter, *Electronics*, April, 1947, pp. 100-105.

⁴ Moore, J. B., and H. A. Moore, I-f Selectivity in Receivers for Commercial Radio Services, *RCA Rev.*, January, 1940, pp. 319-323.

77. Performance of I-f Amplifiers in Receivers for Radar and other Aids to Aviation. In connection with the Hazeltine LANAC (*L*Aminar Navigation and Anti-Collision) system of navigation, i-f amplifiers operating at 60 Mc using the 6AK5 tube¹ have been designed. The performance of these is generally similar to the results obtained in radar i-f amplifiers.² One of the LANAC amplifiers has seven stages and is stagger-tuned so as to have a band width of 12 Mc at 6 db. Despite this wide band width, this amplifier gives a gain of 85 db, or an average of 12 db per stage. In this design the individual couplings are tuned at 55, 60, or 66 Mc.

Wide-band i-f amplifiers operating at 60 Mc have been also designed using negative feedback³ instead of staggered tuning to obtain the required width of band. In one six-stage amplifier using 6AK5's, the tubes are grouped into a first triplet consisting of the first three stages followed by a similar second triplet. In each triplet the inter-stage couplings consist of very closely coupled transformers poled to deliver to the following grid a signal of the same instantaneous sign as on the preceding plate. The negative feedback is then obtained by connecting 33,000-ohm resistors from plate to plate, two such resistors being required in each triplet.

An improved pentode (type 404-A) having less noise and about twice the ratio of transconductance to the total of input and output capacitances in comparison with the 6AK5 is used in the 65-Mc i-f amplifiers in the 4,000-Mc relay system operating between New York and Boston.

Receivers for searching for enemy radar signals included designs⁴ with high i.f.'s. The highest reported was 200 Mc, the amplifier consisting of ten 6AK5 stages with stagger tuning. A band width of 20 Mc at the peak and 24.5 Mc at 6 db was obtained.

78. Volume Control. The function of the volume control is to reproduce the modulation at the desired loudness and preferably to maintain this loudness as the receiver is tuned to different stations of various field strengths or as the field strength of the tuned station varies because of fading. The almost universal general solution to these problems is the use of a.v.c. for securing the proper amount of h-f amplification and the use of a potentiometer in the audio amplifier⁵ for establishing the output volume. Ideally, this method produces a signal at the detector which has the same strength for all stations and fading conditions, so that after the audio volume control is adjusted, no further attention is required.

The manual volume control in a-v-c sets is almost always located between the diode detector and the first audio stage. Usually the one or two diodes for the detection and a triode as first audio amplifier are in a common envelope; e.g., the 12SQ7 and the 14B6. The most common value of the volume control is $\frac{1}{2}$ megohm, with 1 megohm also frequently found. In best designs, the return circuit from the diodes includes first a tweet filter of about 100 $\mu\mu\text{f}$ followed by resistor of 47,000 ohms followed by another 100- $\mu\mu\text{f}$ capacitor, after which the circuit divides into the a-v-c bus and the volume-control potentiometer.

A capacitor is sometimes provided to prevent the flow of d.c. through the volume control, the aim being to increase the life of the volume control before noise develops. A capacitor is generally used in the lead from the arm of the volume control so as to prevent applying the a-v-c bias to the first audio stage.

A special effect as the volume control is reduced can be obtained with a tone-compensated design⁶ in which the apparent strength of the low frequencies is maintained. In the absence of this feature, the low notes disappear and make music sound thin

¹ Mellwain, Knox, Hazeltine Lanac System of Navigation and Collision Prevention, *Proc. Radio Club of America*, February, 1947, pp. 3-28; see specially pp. 21, 25.

² Hopper, A. W., and S. E. Miller, Considerations in the Design of a Radar Intermediate-frequency Amplifier, *Proc. IRE*, November, 1947, pp. 1208-1219.

³ Bode, H. W., "Network Analysis and Feedback Amplifier Design," Van Nostrand, 1945.

⁴ Hulstede, G. E., J. M. Pettit, H. E. Overacker, K. R. Spangenberg, and R. R. Buss, Very-high-frequency Receivers, abstract, *Electronics*, March, 1946, pp. 92, 106. Lebenbaum, M. T., Wideband I-f Amplifiers above 100 Mc, *Electronics*, April, 1947, pp. 138-141.

⁵ Ghiradi, A. A., Volume and Tone Control Resistors, *Service*, September, 1945, pp. 26-27, 30, 41-43.

⁶ D'Orio, P. A., and R. DeCola, Bass Compensation Design Chart, *Electronics*, October, 1937, pp. 37-38.

whenever the volume control is reduced. The circuit in Fig. 16 furnishes a path to ground for the higher audio frequencies which is more effective when the volume control is set at the lower positions; the relative strength of the bass notes is thus increased. Such bass-compensated tone controls have been generally found to afford more pleasing reproduction at low volumes and have therefore been widely used in medium- and high-grade receivers.

Tone compensation for the high notes can also be provided, using the same fixed tap as for low-note compensation. For this purpose, a fixed capacitor is connected in shunt with the upper end of the volume control. This provides a path for the high audio frequencies, delivering them in greater amount at the fixed tap so that they are more effective when the volume is set low.

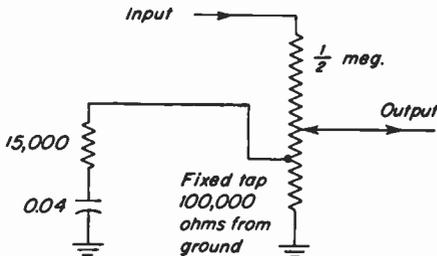


FIG. 16. Bass-compensated volume control.

AUTOMATIC VOLUME CONTROL

Practically all radio receivers now manufactured are provided with a.v.c.¹ The usual practice is to employ gradual cutoff tubes for r-f and i-f amplification and a suitable type of converter so that the bias on all these tubes can be made more negative and the gain thus be de-

creased as the signal strength increases, the bias consisting of d-c voltage developed in a diode detector by the signal itself.

79. A-v-c Figure of Merit. The extent to which an a-v-c system delivers uniform output for various input intensities is indicated by the "a-v-c figure of merit," which is the range in decibels over which the input can vary while producing a change of only 10 db in the output. In particular, with home receivers operating on the broadcast band with terminals for an external antenna, the a-v-c figure of merit is the number of decibels below an input of 20 db ($100,000 \mu\text{v}$) for the 10-db change of output; i.e., the a-v-c figure of merit is the amount of reduction of input below 20 db which will cause the output to decline only 10 db. For loop-type home broadcast-band receivers, the a-v-c figure of merit is the reduction in input below $100,000 \mu\text{v}$ per m (20 db below 1 volt per m) which causes the output to decline 10 db. For automobile receivers and for the h-f ranges of home receivers, the corresponding upper level to be used as reference for the a-v-c figure of merit is 46 db ($5,000 \mu\text{v}$). For home receivers, the a-v-c figure of merit normally lies in the range of 40 to 55 db, being larger for the better receivers.

80. Typical A-v-c Circuits. Figure 17 shows the type of a-v-c circuit now in widest use. It is, of course, a compromise between performance and cost. In it one diode circuit and one load resistor are used to develop both the a-v-c bias and the audio output. A common detector is used for both a-v-c and audio requirements. In operation, the positive peaks of the i-f signal on the diode plate attract electrons from the cathode. These flow through the secondary winding of the i-f transformer, through the 47,000-ohm filter resistor, and through the $\frac{1}{2}$ -megohm volume control, returning to the cathode. When modulation is present, this flow of electrons has both d-c and a-f components. The desired amount of a-f voltage is picked off by the tap and applied to the grid of the triode which serves as the first audio amplifier. The rectified d-c voltage across the 500,000-ohm diode load acts through the 2.2-megohm filter resistor in the a-v-c bus to determine the grid potential of the two h-f tubes which are controlled. As the signal strength at the antenna increases, more signal will be rectified by the diode, and a greater negative bias will be applied to the two controlled tubes. This change of bias reduces the gain of these tubes with the result that the output of the receiver changes by a much smaller ratio than the input.

¹ The original paper on this subject was presented by H. A. Wheeler, Automatic Volume Control for Radio Receiving Sets, *Proc. IRE*, January, 1928, pp. 30-34.

81. Use of Gas Gate with A.V.C. Extensive use is made of a circuit feature called a "gas gate," which offers insurance at negligible expense against a certain type of trouble. Sometimes the converter tube, especially in sets having an r-f stage followed by an untuned coupling, will develop sufficient ionization so that positive ions going

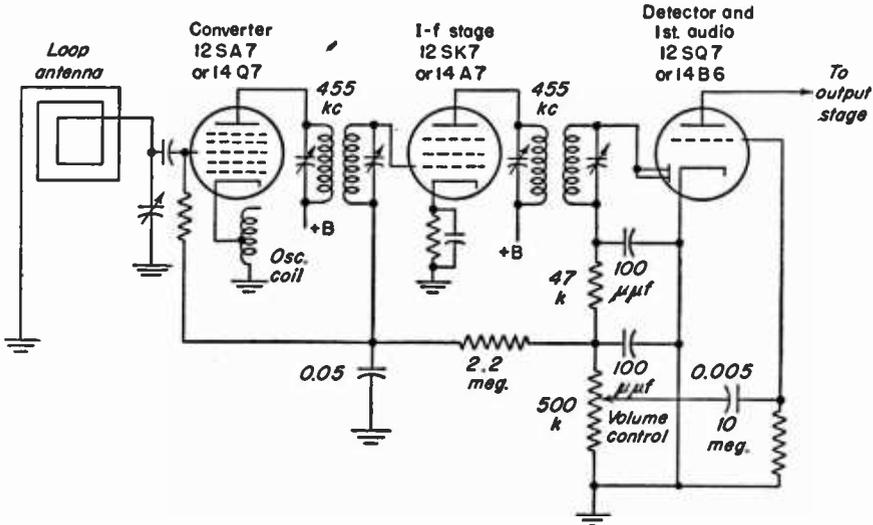


FIG. 17. Typical a-v-c circuit in five-tube receiver.

to the control grid and thence flowing to ground through the high bias resistor will maintain such a large positive d-c drop in the bias resistor (30 or 40 volts) that the tube will draw an excessive sustained plate current and damage will result. This condition can occur with tubes which are normal as indicated by all the usual tests and circuit applications. The gas gate, as shown in Fig. 18, normally employs the second diode plate and connects it to the a-v-c line on the side of the filter resistor which is toward the grid return of the converter stage. Under these circumstances, any positive potential on the grid causes a flow of current through the low forward resistance of the gas-gate diode, which prevents the grid from going appreciably positive. Television engineers will recognize the similarity of this to the reinserting diodes used in studio and receiving equipment.

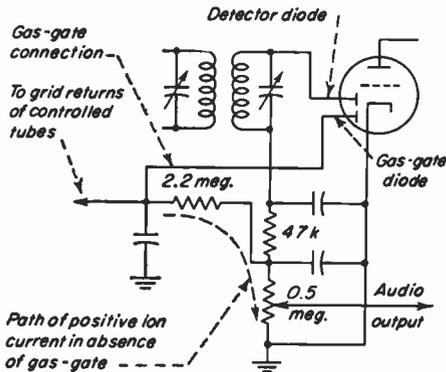


FIG. 18. Essential features of "gas gate" in an a-v-c circuit.

82. Different Selectivity for A-v-c and Audio Detection. Figure 19 shows the use of separate diode plates for the a-v-c and audio detection. This allows the a-v-c diode plate to be supplied with signal from a point in the receiver which is preceded by less selectivity than is used ahead of the audio detector. In fact, the a-v-c diode is supplied from the primary of the last i-f transformer while the audio diode is supplied from the secondary. The advantage of this arrangement is that, when the

receiver is tuned through a station, the bias voltage is held about constant so that the station "naturally" becomes louder owing to the additional selectivity in the audio detector channel and then fades out, permitting easy adjustment of the tuning. Also the large amount of noise heard near each station is reduced. The difference in the signal strength in the two channels should not exceed 20 db. Otherwise the sensitivity in the neighborhood of a strong station may be reduced to the point of preventing the reception of a weak station on the next channel.

In Fig. 19 the a-v-c diode is supplied through a capacitor of $100\ \mu\text{f}$ and has a load resistor of 1 megohm. No filter capacitor can be connected at this diode because it would detune the primary of the transformer. Instead, filtering is provided by a series resistor of 1 megohm and a $0.1\text{-}\mu\text{f}$ capacitor, after which the control bias is delivered to the grid returns of the converter and i-f stages. Additional filtering is provided for the lead which delivers bias to the r-f stage.

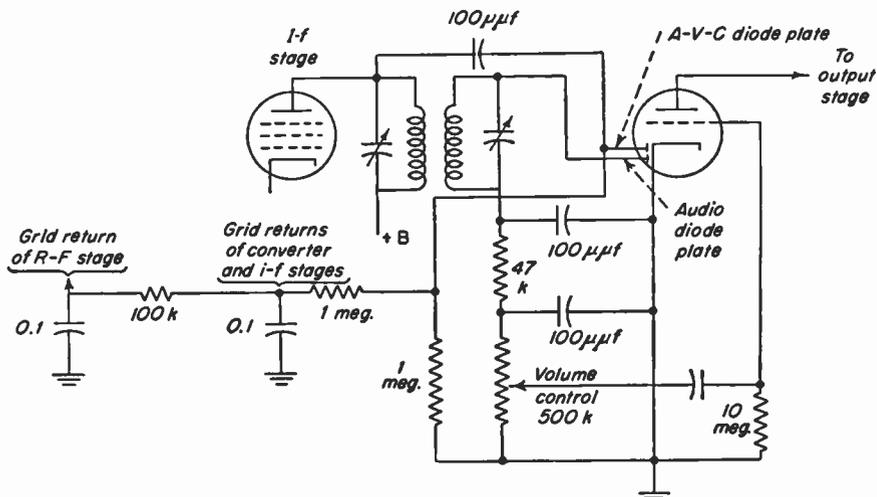


FIG. 19. Use of separate diodes for a-v-c and audio detection.

83. I-f Filtering. It is necessary that the i-f signal at the detector be prevented from proceeding on into the audio amplifier and also from going back into earlier stages of the receiver. The presence of i-f signal in the audio amplifier is objectionable because of (1) the likelihood of producing overloading with consequent audio distortion, and (2) the increased probability of feedback of the i-f fundamental or its harmonics into the r-f or i-f stages. Feedback into these earlier stages can occur directly from the detector in various ways, such as due to poor placing of parts, lack of shielding, or need for increased filtering such as for the a-v-c bias line. Feedback to the r-f circuits will produce whistles (or tweets) on stations having frequencies near the harmonics of the i.f. Feedback of the fundamental into the i-f circuits will produce instability.

To confine the i-f signal the circuit of Fig. 17 includes a filter in the diode plate return consisting of two $100\ \mu\text{f}$ shunt capacitors and a 47,000-ohm series resistor. This filter greatly reduces the amount of i-f signal getting either on the a-v-c line or on the first audio grid. In Fig. 19 a similar filter is provided in the audio circuit and increased general filtering is included in the a-v-c line.

84. Delayed A.V.C. All the a-v-c circuits just described are subject to the objection that a fairly weak signal will develop some negative bias in the load resistor of the a-v-c diode circuit and thereby reduce the r-f amplification to some extent, all under

conditions where the full amplification and the maximum output from the receiver are desired.¹ Various systems have been devised which prevent this effect. In terms of a signal of increasing field strength, these circuits delay the start of a-v-c action until the output has reached a volume as great as desired. The name "delay" is, therefore, with regard to increasing field strength and not with regard to time.

The difference is shown in Fig. 20 where curve *A* is the characteristic which the receiver will have without any a.v.c. The curve *B* is the characteristic with simple a.v.c. The curves *C* and *D* show the characteristics obtained with delay, more stages being controlled in the case of *D* than with *C*. Curve *E* is the ideal, where the output is uniform above the necessary minimum value of signal strength.

85. Biasing Controlled Tubes in Different Amounts. To minimize a type of distortion frequently encountered due to curvature of the characteristic of amplifying tubes, it is desirable as a general principle to proportion the a-v-c bias on the various tubes so as to have less bias on those operating at higher signal levels.

This is a refinement found in the better receivers. The method consists of a tap on the a-v-c load resistor so that, e.g., in the case of two i-f stages, the last i-f stage receives only half the bias applied to the other tubes of the receiver. The cost consists chiefly of one additional resistor and capacitor for filtering the new a-v-c line.

86. Separate A-v-c Amplifier. For very high-grade performance, such as in communication receivers and in point-to-point practice, a separate a-v-c amplifier should be provided. In one form this consists of a separate i-f amplifier followed by a separate diode, so that greater negative biases are obtained and a flatter a-v-c curve is secured. In the other form the a-v-c bias is obtained in the usual way, and the separate a-v-c amplifier consists of a d-c amplifier which increases the magnitude of the bias voltage before application to the controlled tubes.

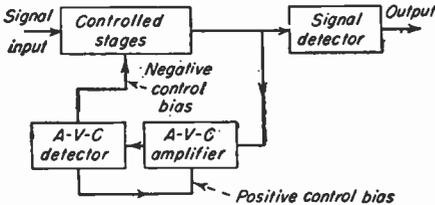


FIG. 21. Use of positive bias in the a-v-c amplifier to obtain a very flat a-v-c characteristic.

The progressive method can sometimes be advantageously employed where a separate i-f amplifier for a-v-c is provided. A very flat a-v-c curve, i.e., a very high a-v-c figure of merit, is obtainable in this way, but care has to be taken to avoid applying too great a bias to the last controlled tube where the signal level may be high. Also, there is a possibility that the a-v-c curve may drop below the horizontal, so that the receiver may give less output for greater input signals, if this method is carelessly applied.

Another a-v-c arrangement that will give a very flat a-v-c curve has been used in

¹ An interesting discussion of remedies for this in point-to-point practice is given by John B. Moore, AGC Noise Considerations in Receiver Design, *Electronics*, May, 1945, pp. 116-118.

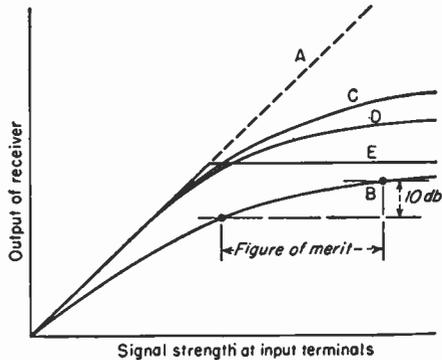


FIG. 20. Various a-v-c characteristic curves.

point-to-point telephone practice¹ and is shown in Fig. 21. In this circuit the a-v-c detector following the a-v-c amplifier provides negative bias for the controlled stages but is also arranged to deliver positive bias to the a-v-c amplifier. The benefit of this positive bias connection can be seen if we suppose that a very large signal is being received and that the output of the set would otherwise be more than desired. The effect of the positive connection is to increase the amplification in the a-v-c amplifier thus making the a-v-c bias more negative and producing the desired reduction in the amplification of the controlled stages. Conversely, when the signal input is very weak, the positive and negative biases delivered by the a-v-c detector are small so that the gain of the a-v-c amplifier, with a small positive relative bias, is small, with the result that the controlled tubes operate at or near full gain.

87. Ratio of A-c Impedance to D-c Resistance. In the proportioning of parts in the diode load circuit for audio detection, especially for high-fidelity receivers, it is important to make the impedance for audio frequencies equal as closely as practical to the d-c resistance. The d-c diode load is usually more or less shunted for audio currents by the elements in the grid circuit of the first audio stage and especially, if the same diode furnishes both a-v-c and audio output, by the filter elements in the a-v-c line. For high-fidelity detection it is desirable that the ratio of a-c to d-c impedance be as nearly unity as possible. Further information on detectors is given in Chap. 10.

88. Gain-time Control and Gated A.V.C. In radar receivers the irregularity and the uncertainty of the return echoes make the employment of a.v.c. undesirable, and two other types of operation are used instead. One of these is gain-time control (GTC), in which after each outgoing pulse the receiver is brought up to full sensitivity at a definite rate so that gain increases as a function of time. In this way the powerful echoes from nearby objects are attenuated while echoes arriving later from distant points find the receiver at full sensitivity and are thereby adequately reproduced despite the weak intensity of these echoes. Gain-time control is generally used, for example, to operate a PPI (plan and position indicator) type of oscilloscope, which is the most widely used means of radar presentation.

Some radars are designed to track a given target, such as an airplane, after they are once trained upon it. For this purpose a.v.c. can be used, provided the particular return echo is isolated and used as the control signal. This isolation is accomplished by means of a "gate" which is a time-sampling process by which, at the particular moment when the echo is arriving, the output of the receiver is routed into the a-v-c diode. Here a suitable control bias is developed as in the usual a-v-c circuit. Manual or automatic means are provided so that the timing of the gate is varied to hold the desired target as it varies in its location. Gated a.v.c. therefore operates to control the gain with regard only to a signal arriving at a definite time following the transmission of the outgoing radar pulse. By means of gated a.v.c. the strength of the particular echo can be held constant, independent of wide variations in the intensity of the echo as received on the antenna.

TUNING INDICATORS

A tuning indicator is a device to show visually when a receiver is accurately tuned to a station. Various forms² of indicators have been used and two of these are still extensively employed. Those now used are (1) a milliammeter in the plate circuit of one or more tubes controlled by a.v.c. and (2) the electron-ray tube, a special form of small cathode-ray tube acting as a voltmeter to indicate when the bias on the a-v-c line reaches its greatest negative value.

89. Meter as Tuning Indicator. The earliest form of tuning indicator, provided in the first a-v-c sets, was a milliammeter showing the current in the controlled tubes. This form is still in general use for communication receivers. Such a meter is often graduated in *S* units to show signal strength according to practice in amateur radio.

¹ Black, D. M., G. Rodman, and W. T. Wintringham, Ultra-short-wave Receiver for the Cape Charles-Norfolk Multiplex Radiotelephone Circuit, *Proc. IRE*, February, 1945, pp. 98-100.

² Schmidt, Fred W., Tuning Indicators, *Radio Maintenance*, December, 1946, pp. 12-15, 20-23.

90. Electron-ray Indicator Tubes. Various types of small cathode-ray tubes specially constructed for use as tuning indicators have been produced. The most widely used is the 6U5/6G5, which produces a sector-shaped pattern with a total opening of 90 deg when there is zero voltage on the control grid, the angle closing gradually to zero as more negative biases are applied to the control grid. The pattern is completely closed at a potential at -8 volts if the plate supply is 100 volts, at -18.5 volts if the supply is 200 volts, and at -22 volts for a supply of 250 volts. A built-in d-c triode amplifier is included in this electron-ray tube. A-v-c voltage is applied to the grid of the amplifier, which produces an amplified positive voltage on a deflecting electrode internally connected to the triode plate.

91. Advantage of Tuning Indicator. Whether or not to provide a tuning indicator in a given receiver is a question which the engineer may have to consider carefully. For an a-m receiver designed to give high-fidelity reproduction, a moderately wide flat-topped response is desirable, and a tuning indicator should be included so that stations can be accurately tuned to the center of this region, where conditions will be best with regard to freedom from interference. For a high-grade f-m receiver, a tuning indicator should also be provided, especially because, when the set is tuned, the program is generally at a medium or low modulation level, and the general user may have little way of knowing when the tuning is accurately centered. With an f-m set, centering the tuning is more important than with an a-m set because, when loud passages of the program occur in the f-m transmission, the frequency swings over a large part of the 150-kc total deviation. The 6AL7GT tuning indicator has three separate deflector electrodes and allows independent display of the i-f tuning and the discriminator tuning, or other arrangements utilizing the versatility of the tube.

92. The 6AL7GT Electron-ray Tube. This tube¹ with three deflecting electrodes has in addition a control grid which can be used to vary the brightness of the display. This tube does not have a built-in amplifier, but its sensitivity amounts to 1 mm height of display per volt. The display consists of two columns of fluorescence each going upward or downward from a central reference point. The two upper portions are separately controlled by two deflecting electrodes, and the two bottom portions are controlled in step by the third electrode. The grid characteristic has a cutoff at -6 volts, giving rise to the possibility that an indicator-squelch feature can be incorporated so as to darken the display between stations.²

SQUELCH AND NOISE-LIMITING CIRCUITS

Two methods for combating noise are squelch and noise limiting. Either or both may be used in a given receiver, and both should be considered in special-purpose designs. Extensive use of both features is made in police receivers, and noise limiters are generally included in receivers for the communication services.

93. Squelch Circuits. Various terms are used to designate squelch circuits, such as quiet automatic volume control (or QAVC), muting systems, and codan (Carrier-Operated Device Anti-Noise).³ The purpose of these circuits is to squelch or quiet the receiver when sufficient signal for satisfactory reception is not present. Where the transmission is not continuous but the receivers must always be ready for a signal, as in police service, a squelch circuit performs the important function of removing the large amount of noise which might otherwise be present between transmissions. In receivers for broadcast reception and other services in which tuning through a wide band is necessary, the squelch feature affords the additional benefit of interstation noise suppression.

Considerable care in the design of a squelch circuit is necessary to avoid "ragged-

¹ Bailey, F. M., An Electron-ray Tuning Indicator for Frequency Modulation, *Proc. IRE*, October, 1947, pp. 1158-1160.

² For other uses of electron-ray tubes, see L. C. Waller, Applications of Visual Indicator Type Tubes, *RCA Rev.*, January, 1937, pp. 111-125; G. O. Thacker and R. Y. Walker, "Magic Eye" Indicators, *Wireless World*, January, 1947, pp. 30-31.

³ An improved codan circuit is described by H. B. Fisher, Remotely Controlled Receiver for Radiotelephone Systems, *Proc. IRE*, April, 1939, pp. 264-269.

edge" effects when the signal fluctuates rapidly between acceptable and nonacceptable values.

A squelch circuit consists essentially of suitable means for (1) assessing the reception conditions and (2) disabling the receiver as long as satisfactory reception is not possible. The assessing is generally on the basis of conditions in the i-f amplifier, and the disabling is usually accomplished by biasing a grid of the audio amplifier beyond cutoff.

The assessing function of a squelch circuit in an f-m system can utilize the rectified limiter grid current, as representative of signal strength, tending to remove the squelch and let the set operate. At the same time a noise signal of audio frequencies above the transmitted voice range can be rectified and used in the direction of applying the squelch. In police service, when the station carrier comes on for a transmission, the limiter voltage in such a set increases, and the rectified noise voltage decreases; since both effects act to remove the squelch, the receiver becomes operative.¹

For the application of the large squelch bias a d-c amplifier called the "squelch tube" is generally used. It is ordinarily a triode. The flow of plate current in the triode causes a voltage drop in the plate circuit, and the connections are so arranged that this voltage drop biases the first audio amplifier beyond cutoff, thus squelching the set. Conversely, when the squelch tube is cut off, the first audio stage operates and signals are received.

Since receiving conditions vary, especially with regard to the amount of man-made interference in the locality, an adjustment is provided to control the level required by the squelch to let the set operate. In this way, when interference is slight, the sensitivity and range of the receiver can be increased. Under the best conditions a signal of 140 db (0.1 μv) is sufficient to open the squelch of a police receiver, and the range of adjustment in a typical set is 140 to 128 db (0.1 to 0.4 μv).

94. Noise Limiters. Various arrangements² of diodes, multigrid tubes, and other

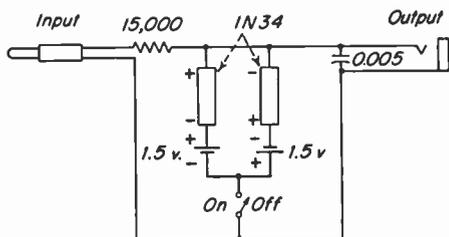


FIG. 22. Simple noise limiter of shunt type for headphone use.

elements are available for limiting the amplitude of noise peaks or disabling the receiver for very short periods while high-amplitude noise is present.

A simple limiter,³ shown in Fig. 22, utilizes two 1N34 germanium crystals. Each crystal is biased by a dry cell so poled that the crystal becomes conductive whenever the amplitude exceeds the 1.5-volt potential of the cell. The maximum a-c signal at the phones is therefore 0.7×1.5 , or about 1 volt rms. This limiter is reported to give good results in the reduction of impulse interference, such as automobile ignition, if care is used in the adjustment of the audio and radio volume controls.

Another limiter, widely used in police receivers and in naval shipboard receivers, is discussed in the Toth reference and shown in Fig. 23. This consists of a diode in series with the outgoing audio lead and having such a polarity that it cannot pass the large negative impulses resulting from the rectification of an intense noise train by the detector diode. The operation may be described briefly by saying that a large noise pulse from the i-f transformer, having peaks of both positive and negative polarities, loses the negative peaks in the regular diode detector, and loses the excess positive peaks, now rectified into a negative peak, in the series limiter diode.

The operation of this limiter depends on time constants and particularly on the fact that the cathode of the series diode requires about 200 times longer to change its potential by a given amount than the plate. A sudden negative peak on the plate therefore

¹ For a squelch circuit of this kind, see Lt. Robt. W. Ehrlich, How FM Links Army Wire Systems, *FM & Television*, April, 1945, pp. 26-31, 85; note especially Figs. 10 and 11 and pp. 30-31.

² Toth, Emerick, Noise and Output Limiters, *Electronics*, November, 1946, pp. 114-119; December, 1946, pp. 120-125.

³ Grammar, George, Noise Limiting in C. W. Reception, *QST*, May, 1946, pp. 13, et seq.

mercy makes the plate greatly negative with respect to its cathode, renders the diode nonconductive, and temporarily open-circuits the audio output connections. By the time the cathode has made a slight start in the negative direction, the noise peak is over and the limiter has served its purpose. For example, the normal potential of the plate may be about -5 volts and the cathode about -7 volts. A noise pulse carrying the plate to -50 volts obviously makes the diode go on open circuit, removing most of the effect of the noise pulse from the output circuit.

In summing up a review of several types of noise limiters, Toth concludes that from the standpoint of simplicity, effectiveness, and low distortion, the series-diode type appears best for voice and m-c-w operation. For c-w operation the full-wave a-f shunt type (such as Fig. 22 with the possible substitution of vacuum diodes for the crystals) appears best because of simplicity, effectiveness, and minimum change of gain below the limiting threshold. Both types can be used in c-w service. The benefit of these limiters is greatest where the noise peaks substantially exceed the desired carrier peak values.¹ If this is not the case, only some very elaborate types will afford much relief.

95. Tone Control. The low cost and utility of a tone control have made it a common feature of broadcast receivers. It can be defined as any control by which the user can vary the amplitude-frequency characteristic of the audio amplifier of the receiver. Its usual form consists of a simple h-f attenuator. To the average listener such a tone control offers the advantages of reducing (1) the worst effects of atmospherics or man-made interference and (2) the high-pitch content of the program if desired.

A survey of 31 models produced during the first two years of the postwar period showed 17 to have tone controls. These consisted of 7 with a series fixed capacitor and variable resistor, 7 with switched arrangements of two to four positions, 2 with independent knobs for bass and treble control, and 1 with multiple push buttons marked for treble, voice, normal, and other types of effects.

The conventional circuit consists of 0.002 to $0.05 \mu\text{f}$ in series with a variable resistor whose maximum value is usually 500,000 ohms. This circuit is connected from a plate or grid in the audio amplifier to ground or to B+, so as to introduce the desired amount of loss for the higher audio frequencies.

The switched arrangements ordinarily consist of two or three elements, e.g., one capacitor and a fixed resistor so connected that the two may be used in series, or the capacitor only may be put in circuit, or neither may be used.

In high-class receivers, there is an increasing trend toward the provision of separate controls for the bass and treble. The treble control can be similar to the series circuits with a variable resistor as already described. Designs including separate bass and treble controls can be made utilizing negative feedback arrangements. Feedback for tone-control uses may be taken from the high side of the voice coil or from the unby-passed cathode of the output stage, and can be brought back so as to introduce a voltage in series with the volume control at its low-potential end. Series or shunt resistors or capacitors associated with the feedback connection make it possible to obtain desired audio characteristics. A tone-compensated volume control can be included in the circuit at slight extra expense.

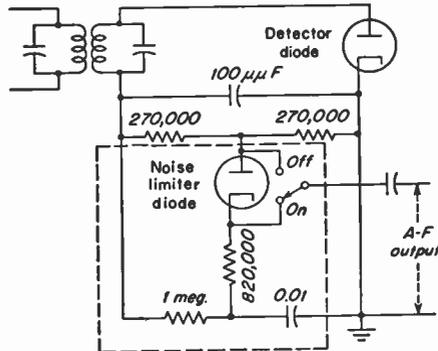


FIG. 23. Simple series-diode type of noise limiter.

¹ Cunningham, W. J., S. J. Goffard, and J. C. R. Licklider. The Influence of Amplitude Limiting and Frequency Selectivity upon the Performance of Radio Receivers in Noise. *Proc. IRE*, October, 1947, pp. 1021-1025.

AUTOMATIC FREQUENCY CONTROL (A.F.C.)

96. Basic Functions. A radio receiver, already tuned approximately to a given signal, can be automatically brought to almost perfect tuning and be maintained there by means of a.f.c.¹ This ability of an a-f-c circuit enables it also to find a signal within a certain frequency range and to capture it, so that the term "signal-seeking circuit" has also been applied. Drift in the frequency of the superheterodyne oscillator due to warmup can thus be compensated, and also tuning errors due to poor manual adjustment or to inaccuracies in the operation of push-button tuners.

An a-f-c circuit consists of two main parts: the *discriminator* and the *control device*. The discriminator is an error detector and determines the direction and amount of the frequency error to be corrected. The output of the discriminator goes to the control device, which is generally a *reactance tube*. The plate circuit of the reactance tube is

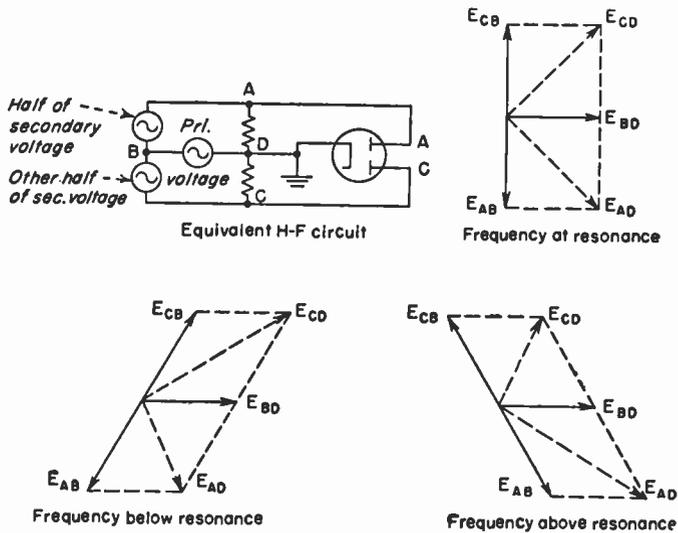


FIG. 24. Action of coupled-circuit discriminator.

made to have a reactive character of controllable amount, so that by connecting it across the oscillator tuned circuit the desired control of the oscillator frequency is obtained. The r-f antenna and interstage circuits are made wide enough not to require correction.

The error found by the discriminator is that existing between the actual signal in the i-f amplifier and the center frequency of this amplifier. The correction made in the oscillator frequency is such that the r-f signal is converted into an i.f. which is much closer to the proper value. Perfect or complete correction with the reactance tube as a control device is not possible because some error is necessary to produce the discriminator output to control the reactance tube. However, this is unimportant in practice because sufficiently complete correction is easily obtainable.

97. Discriminators. The most widely used discriminator² has been the coupled-circuit system, also called the phase-variation type, in which the primary voltage is delivered to the center of the secondary, so that the secondary terminals furnish to two diode plates the primary voltage in the same phase on the two plates and the sec-

¹ Foster, D. E., and S. W. Seeley, *Automatic Tuning, Simplified Circuits and Design Practice*, Proc. IRE, March, 1937, pp. 289-313.

² Roder, Hans, *Theory of the Discriminator Circuit for Automatic Frequency Control*, Proc. IRE, May, 1938, pp. 590-611.

ondary voltage in opposing phases. Both primary and secondary circuits are tuned to the i.f., and this is the frequency which the discriminator uses as its standard of reference. The essential a-c connections and the action are shown in Fig. 24, where it can be seen that (1) at resonance the two diode plates have equal amounts of signal ($E_{CD} = E_{AD}$); (2) below resonance, plate *C* gets more signal (E_{CD}) than *A*; and (3) above resonance, *A* gets more signal (E_{AD}). This action results from the phase relations normally occurring in two inductively coupled tuned circuits. To obtain a d-c output from the discriminator it is only necessary to add the d-c voltages from the two diodes in a series-opposing connection. The resulting single output then has a polarity indicative of the direction of the error and an amount indicative of the extent of the error.

In f-m receivers with a.f.c., a single discriminator can usually be arranged to produce audio signal, a-f-c output, and a-v-c bias. If only a-f-c output is required, there is greater latitude for proportioning the time constants to secure just the kind of a-f-c action desired. Requirements vary depending on whether the receiver is for fixed-frequency or variable-frequency service, the value of the a-v-c time constant, and whether high-fidelity audio output is required.

For further material on discriminators the reader is referred to the general discussion of f-m detectors in Sec. 140 and in Chap. 10.

98. Reactance Tubes. The actual shift of oscillator frequency is made by the reactance tube, whose plate circuit is connected across the tuned circuit of the oscillator. The connections can be arranged for the reactance tube to add either inductive or capacitive reactance in parallel with the oscillator tuned circuit. Inductive reactance in parallel raises the oscillator frequency, while capacitive reactance in parallel lowers it.

To obtain a reactive plate impedance, the a-c plate voltage and plate current must be 90 deg out of phase with each other. The plate voltage comes from the oscillator and can be taken as establishing the reference phase. It is necessary that the current be in quadrature with this. Since the current is determined by the grid voltage, the grid and plate voltages must be in quadrature. An *RC* phase splitter will accomplish this. For example, a small capacitance connected from plate to grid and a small resistor from grid to ground will furnish a circuit from plate to ground which will have an almost pure capacitive reactance. The oscillator voltage will therefore drive a leading current through this circuit. The grid voltage, being the drop in the resistor, will also lead, and therefore the plate current will lead, all with respect to the alternating plate voltage applied by the oscillator. Since the plate current leads the plate voltage, the plate impedance is capacitive, and the desired control of frequency is obtainable. Ideally *C* and *R* in the phase splitter are exceedingly small, and a perfect 90-deg phase shift is obtained. It is desirable that this shift be reasonably near 90 deg so that the plate current contain no substantial component in phase with the voltage. This would represent a loss across the oscillator tuned circuit with the additional objection that it would vary with the amount of frequency correction at the moment.

The magnitude of the capacitance added by the reactance tube in the case just described is approximately $g_m CR$, where g_m is the transconductance of the reactance tube under operating conditions including the bias received from the discriminator. For example, a 6SJ7 reactance tube operating at 1,000 μmhos , with a capacitor of 30 μmf and a resistor of 600 ohms, acts as a capacitor of 18 μmf across the oscillator tuned circuit.

Inductive reactance can be obtained from the reactance tube by exchanging the positions of the two elements, *i.e.*, by using a resistor (with large blocking capacitor in series) from plate to grid and a capacitor from grid to ground. In this case the desired 90-deg relation is most nearly realized with the largest practical values of the resistor and capacitor.

Although a number of broadcast receivers with a.f.c. were made in the late 1930's, including home and automobile models, and gave satisfactory performance, postwar designs have largely omitted a.f.c. and other special features.

99. Use of A.F.C. in Emergency Services. Some use of a.f.c. is now being made in receivers for police, fire, and other emergency services. Figure 25 shows the reactance tube and the oscillator-tripler in the General Electric Model 4-ER-4A-1 receiver, designed for fixed-frequency f-m operation on any channel from 152 to 162 Mc and for installation in either cars or headquarters stations. This receiver is a double super-heterodyne in which the twenty-seventh harmonic of the crystal is used to reduce the signal to the first i.f. of about 17 Mc. The third harmonic of the crystal is used in the second mixer to produce the second i.f. of 2 Mc. The discriminator and the general design are arranged for holding the 2-Mc i-f signal centered on the discriminator characteristic.

The 6AK5 oscillator-tripler, shown in Fig. 25, produces the third harmonic for the second mixer. This harmonic is also used after a further multiplication of nine times as the heterodyne frequency for the first mixer. It is slight variation in this third-har-

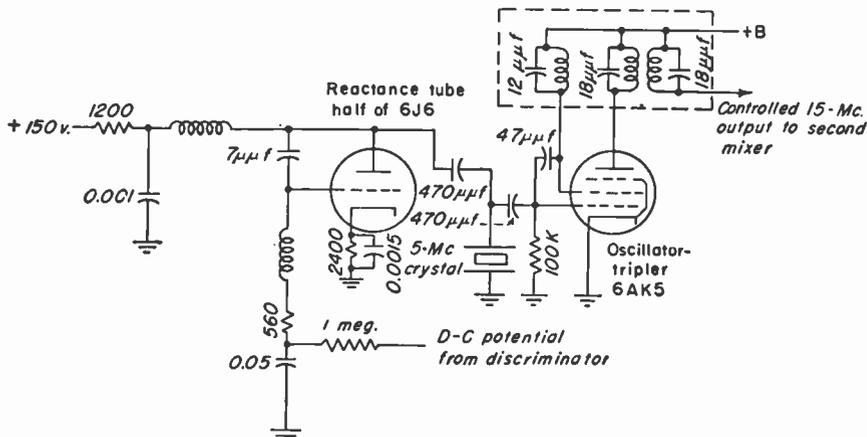


FIG. 25. Use of a.f.c. with crystal oscillator in police receiver.

monic frequency of approximately 15 Mc which automatically keeps the receiver in tune and compensates drifting in either the transmitter or the receiver. Without this a-f-c feature, temperature control of the transmitting and receiving crystals would be necessary. The reactance tube in Fig. 25 is half of a 6J6 triode whose plate resistance is increased by operation at a considerable negative grid bias so as not to damp the crystal excessively.

100. A.F.C. in F-m Broadcast Receivers. In a talk before the Chicago section of the IRE in December, 1946, Lloyd M. Hershey pointed out the advantages of a.f.c. in an f-m broadcast receiver for the 100-Mc band.¹ The control voltage for a.f.c. can be obtained from the f-m detector already in the receiver. It is necessary only to provide the reactance tube and a few resistors and capacitors. An economical solution is to employ a dual triode in which one triode acts as oscillator and the other as reactance tube. The system can be adjusted so as to reduce the tuning error to only $\frac{1}{2}$ to $\frac{1}{10}$ of its previous value.

101. A.F.C. in U-h-f and S-h-f Radar and Relay Receivers. As receiver design has proceeded into the u-h-f and s-h-f ranges, a compelling need for a.f.c. has been found. To accommodate the unavoidable drift of the various oscillators, the alternative has been to increase the band width with the serious objections of reduced gain and increased noise. Radars and relay systems in these frequency ranges, therefore, make

¹ Hershey's discussion is available in his article, Design of Tuners for AM and FM, *Tele-Tech*, August, 1947, pp. 58-59, 95, and in Aram, Hershey, and Hobbs, F-m Reception Problems and Their Solution, *Electronics*, September, 1947, pp. 108-113.

extensive use of a.f.c.¹ One purpose is holding the tuning during warmup, which may be quite important because normal operation may be required very soon after starting the equipment.

Klystrons acting as superhet oscillators can be controlled in frequency by variation of the d-c potential on the reflector electrode. Thus no reactance tube is required, but instead a d-c amplifier to give sufficient values of the voltage to be applied.

The i.f. in receivers for these frequency ranges is generally in the v-h-f band and often is 30 or 60 Mc. A coupled-circuit type of discriminator can be used.

Corrections of as much as 25 Mc may be required and are secured in these services.

102. Mechanical A.F.C. In a mechanical type of a.f.c. the discriminator output polarity determines the direction in which a correcting motor turns, rotating a small air capacitor to the proper point or otherwise making the necessary frequency correction. By gearing the motor down, high accuracy can be obtained. Mechanical a.f.c. has two advantages: (1) the "permanent" nature of the correction, since it does not change in the event of a long fade of the signal—when the signal returns, the tuning is the same and no reestablishment of the correction is required—and (2) perfect correction obtained because the existence of a small error is *not* necessary to maintain the correction. The moving element rotates whenever any appreciable error exists and continues until the error is fully corrected.

The mechanical method of a.f.c. has been used in point-to-point practice² and also in the control of magnetrons³ for radar service.

103. Use of A.F.C. in Double Superhets. The use of a.f.c. in tunable receivers, as for the conventional broadcast band, requires some compromise by the designer to cover from 550 to 1,600 kc (oscillator 1,005 to 2,055 kc). For fixed-frequency service this compromise is not required. Also in double superhets the a-f-c correction can be made on the oscillator supplying the mixer between the two i.f.s, which is a constant-frequency oscillator except for this correction. In these cases very good a-f-c performance can be obtained especially with regard to the frequency intervals at which the system takes hold and lets go.

One prewar broadcast-band design of this kind had i.f.s of 465 and 100 kc. The 365-kc oscillator between the two i.f.'s was varied by the a-f-c action.

SHIELDING AND FILTERING

It is common practice to confine the r-f and i-f circuits in metal containers which provide both electromagnetic and electrostatic shielding. In some instances shielded leads are used to provide the connections to the grids or plates of amplifier tubes, but in general the necessity for such shielding is avoided by so locating these leads that they are sufficiently isolated electrically.

In the design of a new receiver the selection of shielding is mostly a matter of general experience and of cut-and-try work in the laboratory. The presence of a shield can around a coil reduces the inductance and increases the losses, so that allowance for these effects may have to be made in the design. For quite low radio frequencies a fair thickness of shield is necessary, the requirement being a thickness of a few times the depth of penetration⁴ of induced eddy currents at the particular frequency. The depth of penetration is defined as the depth at which the intensity of the induced cur-

¹ Jenks, F. A., Simplified Microwave AFC, *Electronics*, November, 1947, pp. 120-125; and December, 1947, pp. 132-136. Miller, Stewart E., Considerations in the Design of Centimeter-wave Radar Receivers, *Proc. IRE*, April, 1947, pp. 340-351. See section on a.f.c. in R. V. Pound, "Microwave Mixers," McGraw-Hill, 1948.

² Roetken, A. A., A Single-side-band Receiver for Short-wave Telephone Service, *Proc. IRE*, December, 1938, pp. 1455-1465. Rideout, V. G., Automatic Frequency Control of Microwave Oscillators, *Proc. IRE*, August, 1947, pp. 767-771.

³ Hollingsworth, L. M., and R. Dickinson, An Automatic Frequency Control System for Magnetrons with Beacon Applications, 1946, Report PB-55823, Office of Technical Services, Dept. of Commerce, Washington. Helber, Carl A., An Automatic Frequency Control and Frequency Selection System for Magnetrons, 1944, Report PB-55811, Office of Technical Services. Microfilm and photostate are available.

⁴ Wheeler, H. A., Formulas for Skin Effect, *Proc. IRE*, September, 1942, pp. 412-424.

rent has fallen to 37 per cent of its value at the surface. This figure is $1/e = 1/2.72$, where e is the base of natural logarithms. For aluminum at 100 kc the depth of penetration is 0.27 mm (0.010 in.).

Care must be exercised in locating the power transformer and filter reactor on the receiver chassis, since otherwise the electromagnetic field produced by these units may induce an appreciable hum voltage in the a-f circuits. It is desirable to keep these units separated from the a-f circuits as much as possible, and it is frequently necessary to determine experimentally the best location for these components by connecting them into the circuit with flexible leads and orienting them until a position is established which reduces the hum to the desired minimum.

Resistance-capacitance filters are frequently used in the voltage supply leads for the tube electrodes to prevent coupling between points in the system which differ in signal potential and to provide additional filtering for the voltage fluctuations remaining at the output of the B-supply filter. The d-c drop which can be tolerated in a given circuit is frequently a limiting factor in the use of such filters. Filtering can be improved by dividing the maximum resistance value into two or three series parts and by-passing all junctions to ground. When RC filters are used in circuits in which the average current varies during the operation of the receiver, it is essential that the recovery characteristic of the filter be such that the voltage on the electrode can return to its normal value in approximately $\frac{1}{2}$ sec, otherwise noticeable interruptions in the received program will be obtained when sudden changes in the average current occur. This problem is most frequently encountered when RC filters are used in the plate or screen circuits of tubes which are controlled by the a-v-c system.

104. Cabinet. The cabinet must fulfill three requirements:

1. It must house and protect the receiver chassis and loud-speaker mechanism.
2. It must provide sufficient baffle area for the loud-speaker to give the desired l-f response.
3. It must serve as a piece of furniture which will harmonize with the furnishings in the room in which it is to be placed.

Cabinets of wood, plastic, metal, and with leather finish have been found acceptable in the medium and small sizes. Wood is the only material widely used for large cabinets.

Certain proportioning of height and width, in connection with the size, shape, and location of dials and knobs, produces a pleasing appearance, while other proportioning of the various dimensions seems unnatural. A study of these questions has been made.¹

An important cabinet matter is to avoid the occurrence of acoustic feedback which can arise in receivers capable of producing high power output. To avoid this feedback, the tuning capacitor frequently is flexibly mounted, with respect to the chassis, by means of soft rubber washers. The complete chassis in many receivers is also flexibly mounted in the cabinet.

Acoustic feedback is caused by the loud-speaker vibrations being transmitted to the tuning capacitor or some other circuit element which is thereby caused to vibrate sufficiently to detune the receiver intermittently at an a-f rate. If the proper phase relations exist between the loud-speaker vibrations and the variations in signal intensity which result from the vibration of the capacitor plates, sustained oscillations may be produced.

REMOTE CONTROL AND TELEMETERING

Numerous methods are available for the control of radio receivers at a distance and also for the use of radio receivers for controlling other equipment. A third application of remote control is telemetering in which a distant transmitter controls a local receiver so as to furnish an indication or record of conditions in the neighborhood of the transmitter.

¹ Van Dyck, Arthur, *Dynamic Symmetry in Radio Design*, *Proc. IRE*, September, 1932, pp. 1481-1511.

105. Remote Control of Receiver Adjustments. The simplest form of remote control is flexible shafting¹ which may have a length of 2 or 3 ft for the control of an automobile receiver or a length of 15 or 20 ft for the control of an airplane receiver.

S. S. White flexible shafting for this purpose has considerably less angular deflection under torque than the power type used for automobile speedometers. As an example, a remote-control shaft 0.188 in. in diameter (shaft diameter, not diameter of flexible casing) at a torque of 1 lb-in., has a deflection of 3.3 deg per foot of shafting for either direction of torque. The same size of power shafting under the same conditions has a deflection of 15 deg in the winding direction of the outer layer of the shaft and of 32 deg in the unwinding direction.

The "Flexible Shaft Handbook" gives the torsional breaking strength of the various sizes and types of remote-control shafting. For the 0.188 diam this is about 25 lb-in., so that with a safety factor of 2:1, the maximum working load would be about 12 lb-in. Smaller sizes are offered down to 0.117 in. diam, for which the breaking strength is 4.5 lb-in. corresponding to about 2 lb-in. maximum working torque.

Should the torsional deflection be too great, producing too much backlash, which may well be the case with a length greater than a few feet, gearing should be applied so that the shaft turns at a higher velocity and lower torque than would otherwise be the case. A ratio of 14:1 is common and much higher values are sometimes used.

Another widely used method of remote control is by synchros, which are rotatable machines² often called by the trade names Selsyn or Autosyn. One of these is used at each end of the system, with the rotors connected to the same single-phase power-frequency supply line. Each machine has a three-phase stator, the three wires from the one machine being carried the necessary distance and connected to the stator of the machine at the other end. No three-phase power supply is required. If the rotor of either machine is turned, the other rotor follows it closely. Small machines of this kind have been developed by Bendix³ weighing only 4.25 oz each. A system of four of these machines furnishing control and repeat-back services requires only 15 watts from a 400-cycle power source. In this system, the control machines turn at 10 times the angular velocity of the hand crank and are geared down 600:1 to the remote tuning shaft. In this way the shaft is rotated through 180 deg by 30 complete revolutions of the hand crank. The repeat-back system is geared up 1:1.66 so that at the operating point the 180 deg of shaft rotation appears nicely spread out on a 300 deg dial.

Push-button tuning systems of certain types are adaptable to remote control. This is especially the case with motor-operated systems and with systems in which a rotary switch at the receiver progresses cyclically through various positions for the choice of particular stations. (See Sec. 61 *et seq.*)

In communication services, such as at airports, the level of interference is sometimes so high that it is advantageous to locate the receiving antennas and receivers some miles away. In this case, the telephone line used to deliver the audio signal from each receiver can also deliver d-c voltage to control a reactance tube⁴ at the receiver and thus obtain an adjustment of the frequency of the superheterodyne oscillator over a limited range. In c-w telegraph reception, the same method can be used to adjust the beat-frequency oscillator to give the best audio note.

An alternative to the reactance tube, consisting of a coil whose permeability depends on the amount of d.c. in an auxiliary winding, has been described by Polydoroff,⁵ who reported an experimental design with a frequency range from about 200 to 350 kc, *i.e.*, ± 75 kc either side of 275 kc.

Remote control by carrier current on the power wiring, including an arrangement for turning on the receiver at the start, has been described by Kimball.⁶ By means of

¹ "Flexible Shaft Handbook," 2d ed., S. S. White, Industrial Division, 1944.

² Johnson, T. C., Selsyn Design and Application, *Elec. Engr., Trans. Sec.*, October, 1945, pp. 703-708. Holder, L. F., Principles of Selsyn Equipments and Their Operation, *Gen. Elec. Rev.*, 1930, p. 500.

³ McNaney, Joe. T., Remote Tuning Unit for Aircraft Radio, *Tele-Tech.*, March, 1947, pp. 70-72.

⁴ Hannum, E. L., Jr., A Simple Remote Tuning Device for Receivers, *Radio News*, January, 1946, pp. 76, 78, 82. Bard, H. B., Jr., Remote Tuning with Reactance Tubes, *Electronics*, August, 1945, pp. 100-101.

⁵ Polydoroff, W. J., Incremental Permeability Tuning, *Radio*, October, 1944; also published in *Proc. National Electronics Conf.*, 1944, pp. 146-150.

⁶ Kimball, Charles N., A New System of Remote Control, *RCA Rev.*, January, 1938, pp. 303-316.

a high- Q 300-kc circuit in the receiver, a few volts of this signal, added to some 60-cycle voltage from a capacitance voltage divider, enables an 0A4G cold-cathode gas tube to break down and turn on the power switch of the receiver through a relay.

Remote control by means of a special radio wave or special modulation has also been demonstrated. In 1938, Philco introduced the Mystery Control model¹ in which a frequency in the neighborhood of 375 kc was used for change of station, change of volume, and other functions. A small battery-operated transmitter equipped with a telephone-type dial and an internally located loop transmitting antenna constituted the control station, so that no external wires whatever were required. With such a design the control box can be carried anywhere around the house and used at any point to control the radio after it has once been turned on at the main receiver, the maximum range of operation being 30 or 40 ft. The receiver is dual, including a 375-kc control channel in addition to the usual broadcast facilities. The FCC permits the operation of systems of this kind without a license if the field strength does not exceed $15 \mu\text{v}$ per m at a distance equal to the wavelength divided by 2π . At 375 kc (800 m) the field strength therefore must not exceed $15 \mu\text{v}$ per m at a distance of 128 m.

An example of a system in which special modulation is used for remote control was the "alert" receiver² design for communication with air-raid wardens during the war. The receiver is tuned constantly to the frequency of a particular broadcast station and a subaudible modulation is used to connect and disconnect the loud-speaker, 36 cycles being used to turn on the speaker and 24 cycles to turn it off. With this receiver the air-raid wardens could be called instantly during the day or night without having to listen to the regular programs on the station.

106. Use of Receivers for Remote Control of Other Equipment. There has been considerable interest for a number of years in the radio control of model airplanes.³ The typical receiver circuit on the model plane consists of a self-quenching superregenerative one-tube receiver using the RK62 thyratron, whose filament is rated at 1.4 volts and 0.05 amp.

A model boat controlled by radio was demonstrated to many visitors at the Franklin Institute in Philadelphia.⁴ The receiver utilized the RK62 tube and operated on signals consisting of pulses of 60-Mc carrier obtained from an oscillator and a telephone dial. A stepping switch in the model vessel and relays accomplished seven operations including slow speed, fast, reverse, turn to port, and turn to starboard.

Nearly 1,000 water-borne runs of model flying boats were made by remote control using a 117-Mc receiver which with batteries weighed 21 lb.⁵ This receiver included audio filters for the isolation of seven audio and supersonic modulating frequencies from 9,000 to 29,700 cycles, permitting control of elevator, rudder, ailerons, two motor throttles, the flaps, and the ignition. The receiver was a superheterodyne with an 8.8-Mc i.f. The oscillator was crystal-controlled at 9.021 Mc, the twelfth harmonic being delivered to the converter by a chain of a tripler and two doublers. Special features included a battery-compensating tube and the use of a progressive system of a.v.c. A few flights of flying boats were made with this control but most of the work consisted of waterborne runs for the study of hull characteristics in water.

Extensive use of radio for the remote control of airplanes⁶ was made during the late war. These included guided missiles⁷ and also target planes for both aerial and antiaircraft

¹ Herzog, R. G., Mystery Control, *Communications*, October, 1938, pp. 20-21, 29, 31; also *Radio Today*, October, 1938, pp. 30-31, 62-64, and *Service*, October, 1938, front cover and pp. 11-12, 43.

² Deal, H. B., Receiver Control by Transmitted Signal—"Alert" Receiver, *RCA Rev.*, October, 1941, pp. 167-182.

³ Good, Wm. E., Winning the National Radio Control Meet: Details of a Radio-controlled Model Airplane That Has Made Over 100 Successful Flights, *QST*, March, 1940, pp. 24-27, 86, 88.

⁴ West, W. P., Remote Control of a Model Boat, *Electronics*, August, 1940, pp. 19-21. Fitzgerald, G. C., Radio Control for Model Battleships, *Electronics*, January, 1947, pp. 106-111.

⁵ Welge, V., Radio Control of Model Flying Boats, *Proc. IRE*, May, 1947, pp. 526-530.

⁶ Radio Controlled Flight, *Radio News*, January, 1947, pp. 25-27, 137-139. Pilotless Airplane and Missile Control, *Electronics*, April, 1947, pp. 152, 154, 192.

⁷ Ackerman, S. L., and G. Rappaport, Radio Control Systems for Guided Missiles, *Electronics*, December, 1946, pp. 86-91. Selvidge, Harner, Guided Missiles in World War II, *Proc. Radio Club of America*, October, 1946, pp. 3-15.

gunnery.¹ Fighter planes were converted to Ghost Hellcats and used as targets for heavy anti-aircraft guns, the remote-control provisions including radio-controlled ailerons, elevator, rudder, landing gear, flaps, motor, and the independent control of two-wheel brakes. The original Drone, Model TTD-3, was widely used, over 18,000 having been delivered or on order at the end of the war.

A striking use of radio-controlled planes was made during the Bikini atom-bomb tests² by both the Army and Navy who flew such planes through the cloud above the blast for the collection of samples of the atmosphere to be tested for radioactivity. The Navy planes were modified fighters and the Army planes were B17's. The Navy planes took off from the aircraft carrier Shangri-La, flew through the atomic cloud and then were directed to the island of Roi, a distance of 178 miles, as they were too radioactive to permit their return to the carrier. The remote-control equipment included 10 audio channels with filters and synchro motors for carrying out the various functions. In the same way unmanned boats were sent through the contaminated lagoon by radio to collect samples of the sea water while it was still too radioactive to be handled.

Many engineers visualize the time when airplanes can land safely at overcast airports in a completely automatic manner. Demonstrations³ of this kind have been given and it remains a definite possibility for wide-scale use in the future.

107. Telemetering. A telemetering receiver is a local receiver which brings in signals from a distance representing certain measurements and reproduces these on local indicators or records. Such a receiver is a remote-control device in the sense that it operates the indicator or recorder under the control of the remote transmitter.⁴ One telemetering equipment⁵ permits reading a standard aircraft instrument on the ground with very slight modification to the instrument in the plane and operates on the basis of phase angle as the transmitting means. The pickup is therefore a phase-shifting type.

The development of advanced telemetering and remote-control equipment is of special interest in connection with transonic flight,⁶ as the unavoidable dangers are thus transferred from a human pilot to a remote-control robot, and the reports of the test are obtained immediately on the ground by telemetering.

MULTIBAND RECEIVERS

A large number of broadcast receivers produced since about 1935 cover, in addition to the normal broadcast band, one or more ranges in the region from 1.6 to 30 Mc.

108. Receivers with H-f Bands. In most receivers with one or more h-f bands, separate coils are employed in the r-f system for each tuning range. A few receivers use single coils with taps for the various bands. When such coils are utilized, the unused portion of the coil is always short-circuited to avoid absorption of signal at higher frequencies. When separate coils are employed, the coils for two or more of the frequency bands are frequently wound on a single form. The coil windings differ considerably with the frequency range. Wire as small as No. 35 AWG is used in the inductances for the tuning range from 540 to 1,600 kc, while wire as large as No. 22 is used in some of the h-f coils. The turns on the h-f coils are usually spaced to minimize coil losses.

These multiband receivers are provided with a gang switch for simultaneously connecting the coils used for each tuning range to the associated tuning capacitor gang and tubes. On "normal" h-f ranges (as distinct from band-spread ranges) two- or three-gang tuning is used, the circuits being the antenna coupling, the r-f interstage coupling if present, and the oscillator tuning. The band switch connects the proper coil into each circuit for the particular frequency range.

¹ Navy Demonstrates Pilotless Aircraft, *Electronics*, February, 1946, pp. 262, 264, 266.

² Read, Oliver, Drones—Prelude to "Push-button" Warfare? *Radio News*, October, 1946, pp. 25-29, 100, 102, 104.

³ Barbulesco, C. D., Automatic Control of Aircraft, *Elec. Eng.*, March, 1941, pp. 122-126.

⁴ Heeren, V. L., et al., Telemetering from V-2 Rockets, *Electronics*, March, 1947, pp. 100-105, and April, 1947, pp. 124-127. Telemetering in Stratosphere, *Electronics*, March, 1947, p. 142.

⁵ Moore, David W., Jr., Plane-to-ground Radio Telemetering, *Electronics*, November, 1945, pp. 125-127.

⁶ Abbott, Norman A., Electrical Telemetering, *Military Engr.*, October, 1947, pp. 423-429.

109. Control Knobs and Dials. Multiband receivers are sometimes equipped with tuning mechanisms that permit the user to change the drive ratio between the tuning knob and the variable capacitor from a normal value such as 10:1 to finer control such as 50:1. This higher ratio is helpful in tuning the receiver to a h-f station since the band covered in a single h-f tuning range may be over 10 times that covered from 540 to 1,600 kc.

Special tuning dials are necessary on these receivers since a separate scale is required for each tuning range. In some receivers all the scales are visible, regardless of the tuning range being used, and an indicator or lighting arrangement actuated by the range switch knob designates the correct scale. On other receivers only the scale corresponding to the tuning range being used is visible. With this arrangement the dial scales are movable with respect to the dial opening, and the range switch is mechanically connected with the dial scales so that, as the tuning range is switched from one frequency band to another, the proper scale is moved into place.

110. Design of Band-spread Circuits. To the amateur radio operator, the use of band spread on the congested amateur bands is of great value. Likewise, the listener to international broadcasts is greatly assisted by the provision of spread bands for these frequencies. In remote parts of the world where there are colonies of American and European citizens dependent on these frequencies for regular broadcast services, the provision of band spread is quite important.

Various arrangements of capacitors and inductors can be used to spread out a relatively narrow band so that it will cover the entire dial of a receiver.¹

It is often sufficient to provide fixed band-pass circuits for the antenna and inter-stage r-f couplings, so that only the oscillator circuit is varied as the tuning dial is operated. In this case, Foster and Mountjoy² point out that double-tuned antenna coupling gives higher gain and greater uniformity of gain than single-tuned coupling. Typical values for the voltage gain of such a single-tuned circuit covering a complete spread band without adjustment are about 20 db, whereas with a double-tuned circuit values of 24 or 25 db are obtained. In receivers not having an r-f stage, it may therefore be worth while to provide double-tuned antenna couplings on the band-spread positions of the band switch.

An arrangement has been provided in some American models by which two spread bands are accommodated on a single position of the band switch. The lower end of the dial, amounting to about 40 per cent of the total movement, gives the band spreading for one of the bands to be received while the upper 40 per cent gives coverage similarly for the other band. The 20 per cent of the dial between these two parts accommodates in a compressed manner the frequencies lying between these two bands. A special capacitor-plate shape provides these characteristics, the radius changing slowly over the two ends and rapidly in the region between the two bands.

The band-spread ranges on a receiver, to be most useful, must be direct-reading. Since a change of 10 kc at 10 Mc represents only $\frac{1}{1000}$ of 1 per cent, it is evident that the stability requirements on the oscillator tuning are very strict. Special attention must therefore be paid to the effects of temperature, age, voltage supply, and other factors on the oscillator frequency to obtain sufficient stability to satisfy the requirements. If temperature-compensating capacitors are used, care is necessary to locate them on the chassis so that the rate of heat transfer to them is correct to match the oscillator drift.

RADIO-PHONOGRAPH COMBINATIONS

The wide interest of the public in phonograph reproduction since the latter years of the 1930's has made this feature one of the most important aspects of receiver design and production. Great numbers of radio-phonograph combinations have been sold,

¹ Hughes, D. H., *The Design of Band-spread Tuned Circuits for Broadcast Receivers*. *J. IEE*, Part III, March, 1946, pp. 87-96, November, 1946, pp. 459-460.

² Foster, D. E., and G. Mountjoy, *Short-wave Spread-bands in Automobile and Home Receivers*, *Proc. IRE*, May, 1942, pp. 222-227.

including both table model and console types. In addition, separate electric phonographs have been produced in large numbers, and there has been extensive production of wireless record players as attachments to sets not having phonograph provisions. Numerous technical points¹ have arisen in the design and production of this large quantity of equipment, and various special designs have been marketed. The present section summarizes some of the chief engineering data and available methods in this field.

111. Constant-velocity Recording. There are two different fundamental types of recording, constant-velocity and constant-amplitude, a combination of which is generally used.

In constant-velocity recording, if several frequencies are handled in turn by the recording amplifier and are at the same level, each will drive the cutting stylus with the same maximum velocity during each audio cycle. The reproducing stylus will then correspondingly be given the same maximum velocity at the various frequencies when the record is played. Ideally, the magnetic type of cutter produces a constant-velocity recording. This type of recording cannot, however, be used below about 500 cycles because it is accompanied by an increase of amplitude which is inversely proportional to the frequency, with the result that the usual spacing of grooves would be inadequate. The crystal type of cutter will produce a constant-velocity recording if it is operated with a suitable network which introduces increasing attenuation at the higher frequencies.

The intensity of recording on a record can be examined as shown in Fig. 26 by reflecting a light from the surface in the proper manner. The pattern seen is defined as the *optical pattern*, often called a "Christmas tree." It has a uniform width for frequencies recorded with the same velocity.

A constant-velocity recording would in the ideal give a faithful reproduction if played with a magnetic pickup. When played with a crystal pickup, it would give faithful reproduction if a suitable network were used to introduce greater attenuation for the lower frequencies. These points are chiefly of theoretical interest, for the reason that no records are made with constant-velocity characteristics below about 500 cycles.

112. Constant-amplitude Recording. If various frequencies at the same level in a recording amplifier are so transferred to the record that they have the same maximum amplitude on the record, the type of recording and the record itself are characterized as constant-amplitude. This type of recording is universally employed for frequencies below about 500 cycles. A crystal cutter produces a constant-amplitude recording. With a magnetic cutter, a constant-amplitude recording can be obtained by the use of a suitable network with increasing attenuation at lower frequencies.

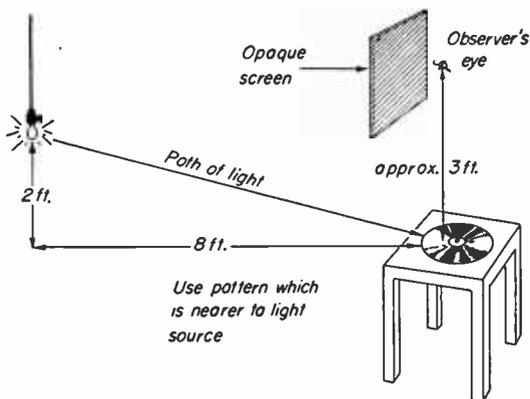


FIG. 26. Arrangement for observing optical pattern of laterally cut phonograph record.

¹ Jorysz, Alfred. *Bibliography of Disc Recording*, *Tele-Tech*, June, 1947, pp. 73-77, 104. Robinson, W. H., *Lateral Recording*, *Communications*, February, 1947, pp. 26, 28; April, 1947, pp. 38-40, 58. Goodell, John D., *The Reproduction of Disc Recording*, *Radio News*, October, 1946, pp. 5-7, 27-29; January, 1947, pp. 3-6, 23; February, 1947, pp. 13-15. Bachman, W. S., *Phonograph Dynamics*, *Electronic Inds.*, July, 1945, pp. 86-89, 124, 128, 190. Roys, H. E., *Intermodulation Distortion Analysis as Applied to Disk Recording and Reproducing Equipment*, *Proc. IRE*, October, 1947, pp. 1149-1152.

In constant-amplitude recording, the maximum velocity is proportional to the frequency since the needle must traverse the given amplitude in less and less time as the period is reduced accompanying the increasing frequency. For this reason the velocity with a constant-amplitude recording doubles each time the frequency is doubled. Therefore for each octave change to a higher frequency the constant-amplitude recording has a 6-db increase of velocity. This corresponds to 24 db greater velocity at 8,000 cycles than at 500 cycles. A proposal has been made¹ that constant-amplitude recording be adopted as standard for general use. This can be considered as a substantial preemphasis on the higher frequencies for the purpose of improving the record by minimizing the amount of scratch noise left after deemphasis in the reproduction. Although this degree of preemphasis is not in general use, preemphasis to the extent of about 9 db at 8,000 cycles is fairly common. In transcription records to be played at broadcast stations, a preemphasis of 18 db at 10,000 cycles in comparison with 500

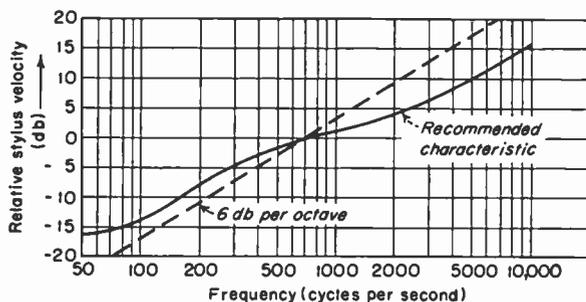


Fig. 27. Velocity characteristic recommended for lateral records by NAB.

cycles is usual for lateral recording in accordance with proposed standards of the National Association of Broadcasters. The curve adopted by NAB for lateral recordings is shown as the solid line in Fig. 27. The dashed line shows a change of 6 db per octave and therefore represents a constant-amplitude characteristic. The horizontal axis represents a constant-velocity characteristic.

A constant-amplitude recording can be faithfully reproduced with a crystal pickup having little or no equalization. It can be faithfully reproduced by a magnetic pickup with the use of a proper network with increasing attenuation at higher frequencies.

113. Transition Frequency. The usual record is recorded with a constant-amplitude characteristic from the lowest frequencies to about 500 cycles, and above this point with a characteristic somewhere between constant-amplitude and constant-velocity. The frequency at which this change from constant-amplitude takes place is defined as the *transition frequency* and is also often called the "crossover" or "turn-over" frequency. It is usually about 500 cycles, but values from 250 to 800 cycles have been used.

114. Tracking Error. The reproducing stylus is not operated vertically against the record but rests back at an angle. It is desirable that this angle be directly above the tangent to the groove, so that the top end of the needle is at the same distance from the center of the record as is the bottom end. Departure from this relation is called "tracking error"² and is usually kept to moderate proportions by placing the pickup cartridge at an angle on the end of the tone arm, or obtaining the equivalent result by a curved shape of the entire tone arm.

115. NAB and RMA Standards. Just before the Second World War, the NAB agreed upon 16 standards³ which have come into wide use at broadcast stations. Some are of interest in connection with records to be played in the home. One provision sets 78.26 rpm as a standard turntable speed, the other standard speed being 33½ rpm. Another gives the standard frequency characteristic for lateral records as shown in Fig. 27.

¹ Duffield, A. W., Improvements in Disc Records through Constant Amplitude Recording, *Communications*, March, 1940, pp. 13-14, 28.

² Bauer, B. B., Tracking Angle in Phonograph Pickups, *Electronics*, March, 1945, pp. 110-115.

³ Smeby, L. C., Recording and Reproducing Standards, *Proc. IRE*, August, 1942, pp. 355-356.

One of the defects to which turntable mechanisms have been subject in the past is flutter (or "wow"), which is the effect of fluctuation in the speed of rotation.¹ NAB standards specify that the maximum instantaneous deviation from the mean speed when making a recording for broadcast-station use shall not exceed ± 0.1 per cent of the mean speed.

Another standardizing activity of NAB consisted in the adoption of standard definitions for numerous terms.² The term "transition frequency" is standardized in preference to "crossover frequency" or "turnover frequency." The term "stylus force" is adopted instead of the erroneous "stylus pressure," which is wrong from a dimensional standpoint. The term "flutter" is chosen in preference to the informal "wow."

RMA issued in June, 1947, Standard REC-105 on disk home recording, which is as follows:

The standard turntable speed of disc home recording equipments shall be 78.26 r.p.m. The second speed of dual-speed equipments shall be $33\frac{1}{3}$ r.p.m.

It shall be standard to locate the record drive pin of disc home recording equipments at a radius of 1 inch from the center of the turntable. The drive pin diameter shall be 0.180-0.185 inch.

It shall be standard to locate the drive pin hole of home recording discs at a radius of 1 inch from the center of the record. The drive pin hole shall have a nominal diameter of $\frac{1}{4}$ inch.

It shall be standard for the center hole of rigid base home recording discs to have a minimum diameter of 0.284 inch.

The standard cutting stylus for disc home recording shall have the following nominal dimensions:

Length.....	$\frac{5}{8}$ in.
Shank diameter.....	0.0625 in.
Face angle.....	87 deg
Heel angle.....	50 deg
Shank flat length.....	$\frac{3}{8}$ in.
Shank flat depth.....	0.010 in.
Maximum tip radius (if any).....	0.002 in.

It shall be standard for the cutting face to be parallel with the stylus axis.

116. Usual Practice in Home Playing of Disks. The usual practice³ in home reproduction is characterized by shellac records, lateral recording, crystal pickup, and approximately 78 rpm.

With regard to such points as dimensions of the groove, size of recording and reproducing styli, and the frequency characteristics, there is a need for standardization.⁴

For good reproduction the reproducing stylus must be larger than the recording stylus; otherwise the reproducing stylus will rattle about in the groove. It is interesting to note that the contact areas involved are so small that the pressure is very high, probably of the order of several tons per square inch. Under this pressure the groove walls yield elastically when the record is played, but it is found that if the reproducing stylus is well polished and properly proportioned and the stylus force is held to 1 oz or less, the deformation is not permanent and the groove is not damaged even by a large number of playings.

The usual crystal pickup of Rochelle salt⁵ is limited in its resistance to temperature and will be damaged if left exposed to a temperature of 52°C (125°F) or higher. This limitation is much less severe with Brush Type PN.

The usual crystal pickup has a capacitance between 700 and 2,000 $\mu\mu\text{f}$, so that the

¹ Furst, Ulrich R., Periodic Variations of Pitch in Sound Reproduction by Phonographs. *Proc. IRE*, November, 1946, pp. 887-895.

² Chinn, Howard A., Glossary of Disk-recording Terms, *Proc. IRE*, November, 1945, pp. 760-763.

³ "Mallory Technical Manual," P. R. Mallory and Co., 1942; see pp. 93-136 on crystal pickups, wireless players, home recorders, and record changers.

⁴ Chinn, Howard A., Disc Recording, *Electronic Inds.*, November, 1946, pp. 64, 66. Beers, G. L., and C. M. Sিনnett, Some Recent Developments in Record Reproducing Systems, *Proc. IRE*, April, 1943, pp. 138-146.

⁵ Dally, Roy, Crystal Phonograph Pickups, *Radio*, September, 1944, pp. 38-40, 46. Cady, W. G., Nature and Uses of Piezoelectricity, *Elec. Eng.*, August, 1947, pp. 758-767

value of 1,500 μf can be taken as typical. With a vertical force of $\frac{3}{4}$ oz (21 g) an output of about 0.5 volt rms is furnished. The capacitance of 1,500 μf corresponds to 1.1 megohms reactance at 100 cycles and 110,000 ohms at 1,000 cycles. It is seen therefore that a high-resistance load is necessary to match the source at low frequencies. This affords a convenient means of adjusting the base reproduction, a load of 2 megohms or more giving a good constant-amplitude fidelity characteristic, while lower resistance values give less output at the low frequencies.

Various selective circuits have been suggested and used¹ for obtaining desired tone response. These usually consist of *RC* networks with occasionally an inductor also employed.

The popularity of the phonograph has been enhanced by the introduction of low-priced record changers² and also by the economy and performance of the rim type of drive in which a large rubber-tired pinion under the turntable furnishes motive power by driving against the turned-down edge of the turntable.

A synchronous motor³ has been proposed as an alternative to the usual induction motor. The suggested design is stated to have very low flutter and to be shielded so as not to induce hum into magnetic pickups.

The usual phonograph combination introduces the audio signal from the pickup into the first audio stage in place of the output from the diode detector. The first audio stage is followed by the output stage, the two furnishing a total gain of about 55 db when the volume control is set at full gain. The volume control is usually a potentiometer at the input of the first audio stage and operates for either radio or phonograph use. Separate phonographs, which have been widely sold, include two audio stages and a rectifier and may also have additional tubes to provide a push-pull output stage and to use a tube inverter in place of the push-pull input transformer.

117. Special Types of Home Reproducers. Various special disk-playing methods have been devised, such as the modulation of a light beam by a mirror attached to the reproducing needle. Another special arrangement⁴ produces a.m. of a 2.5-Mc oscillator by movement of a vane attached to the needle. This design has low mechanical impedance and requires a stylus force of only $\frac{1}{2}$ oz (14 g), so that it produces very little wear of the records. The design has been made and sold under the name "Cobra Tone Arm." Other designs in which the needle produces f.m. of an oscillator have been described.⁵

118. Wireless Phonograph Attachments. One of the standard forms of phonograph equipment is a wireless attachment or phonograph oscillator which produces an a-m wave in the broadcast band for reception by the ordinary broadcast receiver. Many of these designs utilize the l-f end of the broadcast band, being adjustable, for example, from 530 to 730 kc. Others are adjustable similarly at the upper end of the broadcast band, e.g., anywhere from 1,500 to 1,800 kc. These equipments usually consist of a one-tube modulator and a one-tube rectifier. The only external connection is to a 110-volt outlet. This must ordinarily be a.c. owing to the motor power requirement. At the time of installation the frequency of the oscillator is set at an appropriate point well removed from any local broadcast stations. Descriptions of these equipments are given in the Mallory volume to which reference has already been made.

These equipments are permitted by the FCC to operate without a transmitting license on the condition that the field strength must not exceed 15 μv per m at a distance equal to the wavelength divided by 2π .

¹ Bauer, B. B., Crystal Pickup Compensation Circuits, *Electronics*, November, 1945, pp. 128-132. Straede, John W., Audio Response Correction, *Radio-Craft*, September, 1946, pp. 827, 877-878.

² "1947 Automatic Record Changer Manual," Howard W. Sams and Co., Indianapolis. Beitman, M. N., "1945-1947 Post-war Automatic Record Changers and Servicing Information," Supreme Publications, Chicago. "Automatic Record Changers and Recorders," John F. Rider, Inc., 1941.

³ Everett, H. H., A New Synchronous Motor, *Radio News*, February, 1947, pp. 84, 86, 120-122.

⁴ Kalmus, Henry P., Pickup with Low Mechanical Impedance, *Electronics*, January, 1946, pp. 140-146; see also the same author's Improved Modulated-oscillator Pickup, *Electronics*, July, 1946, pp. 182, 184, 186.

⁵ Miessner, Benjamin F., Frequency-modulation Phonograph Pickup, *Electronics*, November, 1944, pp. 132-133. Beers, G. L., and C. M. Sinnott, Some Recent Developments in Record Reproducing Systems, *Proc. IRE*, April, 1943, pp. 138-146.

119. Frequency Records for Testing. Various firms have made frequency records which are of much value in experimental development. The following are lateral records operating at 78 rpm:

Number	Producer	Description
78-1 10003-M 84522 D-61-A	Audiotone Columbia RCA Universal Microphone	250-cycle transition frequency 300-cycle transition frequency 500-cycle transition frequency Constant-velocity recording of 50-200 cycles at level of 7 db; then 200-500 cycles at 14 db; then 500-10,000 at 21 db; then 1,000 cycles in 2-db steps from 8 to 18 db; then 400 cycles at 18 db. All levels are with respect to an arbitrary reference
1000-D	Clarkstan	Constant-velocity, 5,000-15,000 cycles

The following are lateral records operating at 33 $\frac{1}{3}$ rpm:

Number	Producer	Description
2485 TRL-100	RCA Western Elec.	500-cycle transition frequency; has orthoacoustic characteristic, very similar to NAB characteristic 250-cycle transition frequency

120. Dynamic Noise Suppression. A circuit known as a "dynamic noise suppressor" has been introduced by Scott to improve the reproduction of noisy phonograph records¹ on the following basis. When music is at a low level, the lowest and highest pitches are inaudible. There is therefore no harm done to the quality of reproduction if the system introduces high attenuation for these components. Under these circumstances noise present in the same frequency regions is removed, and the reproduction may sound quieter. This noise suppressor is, therefore, a device which controls the frequency band in the reproduction and varies it automatically in accordance with the volume and frequency characteristics of the music. The frequency band is widened during loud passages and contracted during weak passages, the action being independent at the bass and treble ends of the audio range.

The noise to be reduced consists of scratch in the treble region and motor rumble in the bass region. These noise levels, except for a small portion of the time, may be more intense than the musical components existing in the same frequency regions. It is possible to control each end of the spectrum by means of signals present in a region nearer to the center of the audio range. Control in this way resides effectively in the signal rather than in the noise. With music this is possible, since high harmonics exist only when there are fundamentals of somewhat lower frequency which can be used to operate the control. Similarly at the bass end, whenever deep fundamentals are present they are accompanied by harmonics of appreciable amplitude which can be used to control the cutoff.

A model of this suppressor designed for use in broadcast stations has 13 tubes. It provides 20 db of scratch reduction and about 15 db of rumble reduction. Models suitable for home use add two or three tubes to the normal receiver complement.

121. Home Recorders. There is an increasing interest in the home recording² of favorite radio programs as well as the voices of children, members of the family, etc. Considerable equipment for this purpose has been sold and is described in the Mallory volume already referred to. The recording function in such an equipment requires that it have a stronger motor than that provided with turntables for reproduction only. Separate cutting and reproducing heads are provided, as the angle, size, and force for the two styli are different.

¹ Scott, H. H., Dynamic Noise Suppressor, *Electronics*, December, 1947, pp. 96-101.
² "How to Make Good Recordings," Audio Devices, Inc., 1945.

122. Home Recording with Magnetic Mediums. Increasing interest is being shown in magnetic recording and reproducing in the home.¹ This is because of the following advantages: (1) longer playing time, (2) negligible wear in repeated playing, (3) adaptability to erasure and reuse, and (4) compact size for storage.

Supersonic magnetic "bias" of 30 to 40 kc is used to improve the linearity in recording,² and also for erasing. With wire of 0.004 in. diameter as the medium the speed is about 24 in. per sec. Nonmagnetic tape coated with magnetic particles can be used at rates of travel as slow as about 8 in. per sec.³

VOLUME EXPANSION AND COMPRESSION

It is possible in an amplifier to vary the gain in such a way that the greater the input, the greater the gain. Such an action is called "volume expansion" because the range of volume applied to the amplifier is increased or expanded. Conversely, the gain may be varied in the opposite direction to produce volume compression.

123. Use in Point-to-point Radiotelephony. Volume compression and expansion are a well-established method of improving the signal-to-noise ratio in point-to-point radiotelephone practice. In this field the term "comparator" is used to designate the compressor and expander equipment at the two ends of the radio link. The compressing and expanding operations are performed at audio frequency and require the use of suitable time constants and other refinements.⁴

124. Compression in Broadcasting and in Phonograph Recording. In broadcast practice symphonic music is compressed by strengthening the weak passages with a manual volume control and reducing the loud passages by either manual or automatic means. Such compression is considered necessary to keep up the average percentage of modulation of the station to deliver a usable signal in the outer parts of the service area, and at the same time avoid overmodulation on the strong passages. In recording phonograph records, compression is similarly used to make the music conform to the limitations of the medium. Both the radio and record reproduction are, of course, normally carried out without any compensating expansion.

There is a question as to whether the listener under typical conditions desires the reproduction of a classical program at the full volume range. To have such a true rendition of the music, the listener must be in a quiet place so that he can hear the pianissimo passages without difficulty, and he must also be in such a place that the fortissimo passages will not cause trouble.

In some situations, *e.g.*, in receivers operating in automobiles (or receivers for speech in locomotive cabs), it may be desirable to introduce compression rather than expansion so that the volume may be held near a level sufficient to be heard easily over the noise and yet not rise to objectionable loudness during high-modulation periods.

125. General Methods. Equipment with an expansion feature has been offered to the public in a few models during past years. These have been intended especially for phonograph reproduction, where the operation is facilitated by the fact that the compression is more systematic than with broadcast radio programs.

One de luxe method of obtaining compression and expansion in a system is to have a separate pilot control channel, which transmits a signal at all times representing the amount of compression at the transmitter and controlling the compensating expansion at the receiver. For broadcast service a 20-cycle tone modulation has been proposed

¹ Pugsley, D. W., *Wire Recording*, *Elec. Eng.*, July, 1946, pp. 316-321. Holmes, L. C., *Some Factors Influencing the Choice of a Medium for Magnetic Recording*, *J. Acoust. Soc. Am.*, May, 1947, pp. 395-403. Kornei, O., *Frequency Response of Magnetic Recording*, *Electronics*, August, 1947, pp. 124-128.

² Holmes, L. C., and Donald L. Clark, *Supersonic Bias for Magnetic Recording*, *Electronics*, July, 1945, pp. 126-136.

³ A 133-page general account of magnetic recording developments in Germany, prepared by R. H. Ranger, is available in mimeograph form from Office of Technical Services, Dept. of Commerce, Washington. Ask for publication PB-79558. This and 41 other German and United States reports of war-work on magnetic recording are included in a free price list issued by the Office of Technical Services in October, 1947.

⁴ For further details, see C. W. Carter, Jr., A. C. Dickieson and D. Mitchell, *Application of Compressors to Telephone Circuits*, *Elec. Eng., Trans. Suppl.*, December, 1946, pp. 1079-1085.

for this purpose, but there appears very little likelihood that any steps along this line will be taken.

In the absence of a pilot control channel, expanders for broadcast and phonograph use must have different time constants for their operate and release actions. In particular, the time constant for taking hold must be quite short, such as 10 msec. The release time must be much greater, such as $\frac{1}{2}$ to 1 sec. The amount of expansion can reasonably be about 10 db at the loudest passages.

126. Expansion by Tube Characteristics. One method of obtaining expansion is by control of the grid bias of an amplifying tube.¹ Such control is usually made progressive, or forward acting; i.e., signal is rectified and the resulting bias used to vary the gain of a tube farther along in the system. Either diode tubes or germanium crystals, such as the 1N34, can be used for the development of bias for expansion.²

An alternative method of using a tube for expansion is to connect the plate circuit of a triode so that its resistance is part of a gain-control potentiometer in the amplifier. The 6C5G triode has been used in series with a fixed 50,000-ohm resistor for this type of operation.

127. Expansion by Nonlinear Resistance. The wide variation of resistance of lamp filaments with current or voltage has been used to furnish volume expansion. However, thermistors,³ (resistors whose value varies with temperature) are more useful for this purpose. In a typical case the resistance is 50,000 ohms at a very low electrical level but declines to only 800 ohms when dissipating 100 mw. These thermistors can be used in shunt across the circuit to produce compression or in series in the circuit to produce expansion. By arrangement in a bridge manner, d.c. can be passed through them to produce further control of the amount of expansion or compression, the d.c. being developed from the signal itself.⁴

A striking application and extension of compression and expansion technique was demonstrated in 1941 by Harvey Fletcher and others of the Bell Telephone Laboratories.⁵ This included the feature of manual control of the expansion by a musical conductor, who thereby was effectively given an orchestra of a thousand pieces available at the turn of a knob. This possibility of "enhancement" of the original volume range affords very impressive musical effects. A three-dimensional feature was also shown by the use of three channels with speakers at the two sides and center of the stage. The volume range of the "stereophonic" film was 50 db which was increased to 80 db by the use of 30-db normal compression and expansion. The enhancement was obtained by further expansion when desired. This system included three pilot channels, one for the control of the volume of each of the loud-speakers on the stage, the control frequencies being 1,260, 3,780, and 6,300 cycles. The single film included three sound tracks for the three loud-speakers and a fourth track on which the three pilot frequencies were recorded. In the reproduction the three pilot frequencies were separated from each other by audio filters and each then used to control the volume of its assigned channel.

Enhancement has a musical value in an additional way on account of the fact that a musician finds difficulty in producing good notes at very low intensity. In a system with enhancement facilities, the enhancement can be used "in reverse" so that such passages can be played at medium or low-medium intensity and then be rendered at the desired very low intensity with perfect tone quality.

¹ Buts, A. Nelson, Jr., Surgeless Volume Expander, *Electronics*, September, 1946, pp. 140, 142. Ehrlich, R. W., Volume Expander Design, *Electronics*, December, 1945, pp. 124-127.

² Moses, R. C., A Volume Expander Compressor Preamplifier, *Radio News*, June, 1946, pp. 32-34, 146-149.

³ Batcher, Ralph R., Thermistors in Electronic Circuits, *Electronic Inds.*, January, 1945, pp. 76-80. Becker, J. A., C. B. Green, and G. L. Pearson, Properties and Uses of Thermistors—Thermally Sensitive Resistors, *Bell System Tech. J.*, January, 1947, pp. 170-212.

⁴ Bennett, W. R., and S. Doba, Vario-Loesser Circuits, *Elec. Eng., Trans. Sec.*, January, 1941, pp. 17-22.

⁵ The equipment and methods are described in seven papers in *J. Soc. Motion Picture Engrs.*, October, 1941, pp. 331-426. These papers are also available gratis as *Monograph B-1327* from Bell Telephone Labs, New York.

In this demonstration the thermistor type of element was used for both the compression in making the film and the expansion in reproduction.

As another means of improving reproduction the stereophonic demonstration included a preemphasis on the higher frequencies and corresponding deemphasis in the reproduction. These amounted to 15 db at 8,000 cycles in comparison with 500 cycles. With regard to the volume expansion, one of the papers listed in the reference mentions that the provision of delay is desirable but was not ready in time for the demonstration. Such a delay slows down the signal in the main transmission path while the gain of the amplifier ahead is being adjusted to the proper point. Delay can be used in systems with or without a control channel. In the case of a control channel, a very easy method of obtaining the delay is to use a separate microphone for picking up the sound to be converted into control signals; it is then only necessary to locate this microphone and the regular microphone at suitable distances to obtain the desired amount of delay.

Compression as a feature of a radio receiver is very seldom used. However, it may be desirable to incorporate it in receivers to be used in noisy situations, such as in automobiles and locomotives. Even in the home when listening in the next room, the usual range of volume may be too great for the best practical results, the faint passages being inaudible, or the loud passages objectionably strong if the volume is turned up sufficient to hear the faint ones. Another instance where compression may be considered is with a home recorder, either of the disk or wire type, as a means of assisting the nontechnical user in obtaining satisfactory recordings.

A-C-D-C RECEIVERS

The a-c-d-c type of power supply is popular because of (1) the economy of not requiring a power transformer and (2) the adaptability to either a-c or d-c power supplies.

128. Essential A-c-D-c Features. In the a-c-d-c power system the heaters of the tubes are operated in a series circuit connected across the line, and the plate circuit is operated as an additional load circuit with a half-wave rectifier in series. Figure 28 gives a typical circuit of an a-c-d-c receiver, and it may be noted that the power plug is marked with + and - indications. This polarity must be observed when the set is operated from a d-c outlet. The set is inoperative when plugged into a d-c outlet with the opposite polarity because the tubes have no plate supply. When the set is operated on a.c., heater current flows throughout the a-c cycle, but plate current flows only when the a-c wave has the proper polarity and is near its peak value. The 40- μ f capacitor connected from the cathode of the 35Z5GT rectifier to B- maintains most of its charge during the power-frequency cycle so that it is only at positive peaks of the a-c wave that the rectifier passes current and this capacitor assumes its maximum charge. The action is similar to peak detection.

On account of this pulseline flow of plate current from the a-c line, an a-c-d-c receiver may not operate satisfactorily when an external series resistor is used to adapt a 110-volt set to a higher a-c socket voltage, such as 220 or 250 volts. Separate resistors for the A and B supplies, or other special arrangements, may be necessary.¹

Selenium rectifiers are now available as a substitute for the tube type and have greater compactness and longer life.

Capacitor-input filters are always used in a-c-d-c receivers. If the first capacitor in a set with a tube rectifier materially exceeds 40 μ f, the set will be subject to failure whenever the user turns it off and then quickly turns it on again. Under these conditions the filter capacitor will quickly discharge, the rectifier cathode will remain hot for the short time involved, and the result will be that, when the set is turned on again, the tube current limit of 500 ma may be greatly exceeded by the rush of current into the capacitor, ruining the rectifier tube and possibly other parts of the receiver. The design in Fig. 28 includes a 22-ohm resistor in series with the rectifier plate as a protec-

¹ This condition is often encountered in England and is discussed in A.C./D.C. Voltage Dropping, *Wireless World*, July, 1946, pp. 236-237.

tion against this effect. This resistor often has a fuselike type of construction so that, if overloaded owing to a short circuit anywhere on the plate supply system, it will burn out without damaging the rectifier tube. A study of these matters made by Tung-Sol led to the recommendation of the following resistance values for protection of the rectifier tube:

Capacitance of First Filter Capacitor, μ f	Series Resistance, Ohms
60	28
80	45
100	65
120	85

Servicemen have been advised¹ to install a 50-ohm resistor in all a-c-d-c sets not already equipped for the protection of the power rectifier. This gives sufficient protection unless a very large first filter capacitor is used.

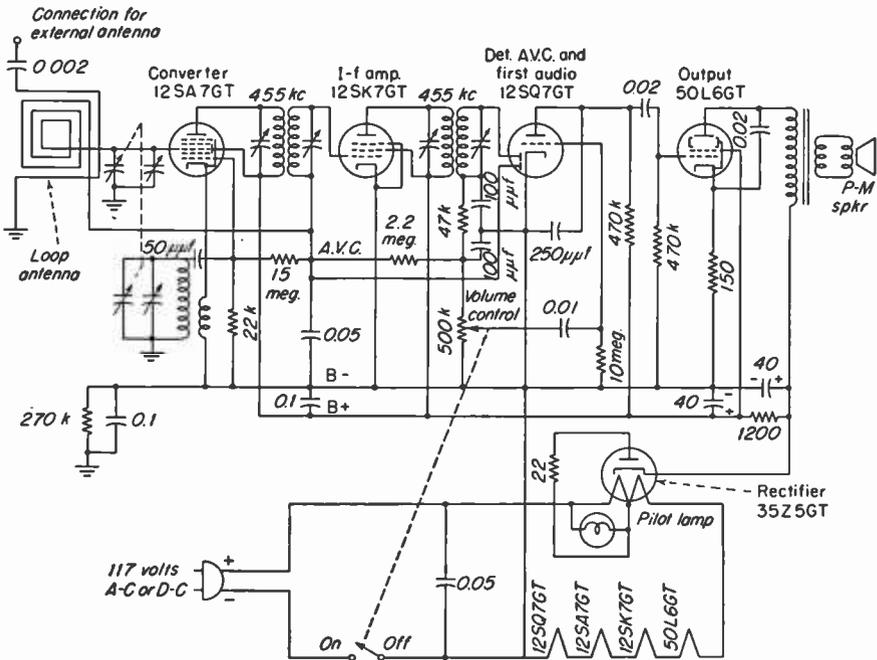


FIG. 28. Typical a-c-d-c receiver.

Higher plate voltage in an a-c-d-c set when operating on a-c supply can be obtained by the use of a voltage-doubler rectifier circuit.

Since a-c-d-c sets are usually housed in small cabinets of limited acoustic properties, various suggestions have been made for improving the audio quality and in particular for improving the apparent amount of bass in the sound. One arrangement advanced by F. H. Shepard, Jr., is to introduce a small amount of distortion at all volume levels in the audio amplifier, there being about 5 per cent of third harmonic always present and a small amount of fifth harmonic. In this way the output-input characteristic is changed from a straight line into a slightly S-shaped curve. The resulting distortion is stated to give the listener the impression that the set has fair output at quite low

¹ "Mallory Technical Manual," P. R. Mallory & Co., Indianapolis. See pp. 48-62.

audio frequencies. This feature was demonstrated to the Radio Club of America in September, 1941, and has been described briefly in the literature.¹

129. Printed Circuits. The proximity fuse development of the Second World War included a new wiring technique in which printed conducting lines on insulating material serve as conductors, and very compact arrangements of impedance elements are made. Various descriptions of this technique have appeared in the literature.² The full import of this development has not yet been determined.

130. Reflex Principle. Although little used in the United States, the reflex principle is one of the standard methods available to the radio design engineer. A reflex circuit is defined in the IRE Standards of 1938 as one through which a signal wave passes both before and after a change of frequency. Since such a circuit has currents of two different frequencies, it has two inputs and two outputs, and filtering is necessary to separate the two outputs.

The field of usefulness is where power levels are low and the number of tubes is restricted by economic or technical factors. European broadcast receivers have often included a reflexed tube serving as both an i-f and a-f amplifier. Both pentodes and converter tubes have been used in this way. In the case of the converter tubes, the mixing feature is not required, but the two input grids offer convenient points for applying the two frequencies. Another possibility is to amplify both the r-f signal as received and the i-f signal, but such designs are likely to have poor performance for radio frequencies which are harmonics of the i.f.

A tube operating in a reflex circuit is subjected to the sum of the peak voltages and currents at both operating frequencies, so that the power levels must not be too high. If care in this regard is not taken, distortion will occur; in particular, tubes handling i-f and a-f signals may give poor audio quality. Besides keeping the power levels low, points worth some consideration are increasing the plate-supply voltage within the rating of the tube, or choosing a larger tube. The advantage obtainable by reflexing has been given by one writer³ in a statement that the sum of the gains at the two frequencies can be within 6 db of the corresponding sum for the tube handling each frequency alone. The power limitation will probably make it undesirable to apply a.v.c. to a reflexed stage; however, if the stage handles i-f and a-f signals and a.v.c. is applied, the control of the a-f gain constitutes forward-acting a.v.c., so that a very flat a-v-c curve should be obtained.

F-M RECEIVERS

Frequency modulation is extensively used for such purposes as police service around 35 Mc, broadcasting around 100 Mc, automobile and railroad telephone around 155 Mc, and the sound channels of television.⁴

131. Modulation Index. The ratio of maximum frequency swing to one side divided by the highest frequency component in the modulating signal is called the *modulation index*. In police service the deviation is usually 15 kc to each side, and the highest a.f. transmitted is 3 kc, giving a modulation index of 5. For present-day broadcast practice the corresponding figures are 75 and 15 kc, again giving a modulation index of 5.

On certain segments of the amateur bands the FCC has authorized the use of "narrow-band" f.m. which is subject to the requirement that "the peak deviation is limited

¹ Improved Bass for Small Radios, *Electronics*, July, 1945, pp. 224, *et seq.*

² McDonald, W. W., Stamped Wiring, *Electronics*, June, 1947, pp. 82-85. Murray, A. F., Evolution of Printed Circuits for Miniature Tubes, *Tele-Tech*, June, 1947, pp. 58-61, 111-112; also Present Status of Printed Circuit Technics, *Tele-Tech*, December, 1947, pp. 29-33. Brewster, F. R., British Metal-spray Circuits, *Electronics*, July, 1947, pp. 158 and 160. Brunetti, Cleo, and R. W. Curtis, Printed Circuit Techniques, *Proc. IRE*, January, 1948, pp. 121-161.

³ Smith, Langford, "Radiotron Designer's Handbook," published in Australia in 1940; sold in United States by RCA Manufacturing Company, Harrison, N.J. Chapter 21, pp. 183-184, treats briefly of reflex methods.

⁴ A general discussion and an extensive bibliography on f.m. are given by S. W. Seeley, Frequency Modulation, *RCA Rev.*, April, 1941, pp. 468-480. Hund, August, "Frequency Modulation," McGraw-Hill, 1942. Kiver, Milton S., "F-M Simplified," Van Nostrand, 1947.

to a value equal to or less than the maximum modulation frequency." This corresponds to a modulation index of unity or less.

132. **Transmitter Specifications Affecting F-m Broadcast Receivers.** For f-m broadcasting the FCC "Standards of Good Engineering Practice" establishes 88 to 108 Mc as the f-m broadcast band, divided into 100 channels of 200 kc each. These channels are arbitrarily numbered from 201 through 300. The accompanying table gives the frequencies for the even-numbered channels. The odd-numbered ones can readily be obtained by adding or subtracting 0.2 Mc, which is the spacing between channels. The frequency given is the center of each channel, the total width extending to 100 kc below and 100 kc above the stated figure.

F-m Channel Numbers

No.	Mc	No.	Mc	No.	Mc	No.	Mc
202	88.3	228	93.5	252	98.3	278	103.5
204	88.7	230	93.9	254	98.7	280	103.9
206	89.1	232	94.3	256	99.1	282	104.3
208	89.5	234	94.7	258	99.5	284	104.7
210	89.9	236	95.1	260	99.9	286	105.1
212	90.3	238	95.5	262	100.3	288	105.5
214	90.7	240	95.9	264	100.7	290	105.9
216	91.1	242	96.3	266	101.1	292	106.3
218	91.5	244	96.7	268	101.5	294	106.7
220	91.9	246	97.1	270	101.9	296	107.1
222	92.3	248	97.5	272	102.3	298	107.5
224	92.7	250	97.9	274	102.7	300	107.9
226	93.1			276	103.1		

The FCC adopted in 1947 a rule that the normal spacing between stations in a given locality will be 800 kc to avoid interference difficulties found in tests in Syracuse and also to permit the construction of economical receivers having a limited number of tuned circuits for selectivity.

Modulation of 100 per cent in f-m practice does not have a unique definition such as it has in a-m practice. The FCC has standardized this for f-m broadcasting as ± 75 kc. On this basis 30 per cent modulation, which is often used for testing, is a deviation of ± 22.5 kc.

In FCC "Standards of Good Engineering Practice," the signal intensities stated as necessary for satisfactory reception are 1,000 μv per m for city, business, or factory sections, and 50 μv per m for rural sections. Objectionable interference is considered to exist for stations on the same frequency if the undesired field strength exceeds $\frac{1}{10}$ of the desired. For stations separated by 200 kc, the corresponding figure is $\frac{1}{2}$. Values for 400 and 600 kc had not been established by mid-1947. For 800 kc and greater it was stated that no restriction applies except that stations in a given area will normally not be assigned frequencies differing by 10.7 Mc, since this is the usual i.f. for f-m receivers.

The audio characteristic of f-m broadcast transmitters must extend from 50 to 15,000 cycles, with preemphasis on the upper audio frequencies in accordance with Sec. 133. The maximum audio harmonic distortion is specified as 3.5 per cent total harmonics for any fundamental frequency between 50 and 100 cycles, 2.5 per cent for any fundamental from 100 to 7,500 cycles, and 3.0 per cent for any fundamental from 7,500 to 15,000 cycles. The f-m noise level in the transmitter output must be at least 60 db below full modulation; i.e., the noise deviation must not exceed 75,000/1,000, or 75 cycles. The a-m noise level must be at least 50 db down, which for a 1-kw station means less than 10 mw.

The FCC also divides broadcast transmitters into classes A and B. Class A stations are local with a power of 0.1 to 1 kw and normally an antenna 250 ft high. Class B stations are regional with a power of 1 to 20 kw and an antenna height of 300 to 500 ft.

133. Preemphasis of Upper Audio Frequencies. Since most of the energy in speech and music is contained in the lower audio frequencies, it is possible to reduce the noise in a system by reinforcing the medium and high audio frequencies at the transmitter and then correspondingly attenuating these regions at the receiver. In this way the power output of the transmitter for these frequencies is increased, requiring a slight reduction of power for the lower frequencies, while a large reduction of noise is obtained by means of the compensating attenuation at the receiver. For this reason the FCC specifies that "preemphasis shall be employed in accordance with the impedance-frequency characteristic of a series inductance-resistance circuit having a time constant of 75 μsec ." The significance of this statement, which specifies preemphasis in microseconds, can be visualized by assuming that the inductance and resistance are connected in series in the plate circuit of a constant-current audio source, such as a pen-

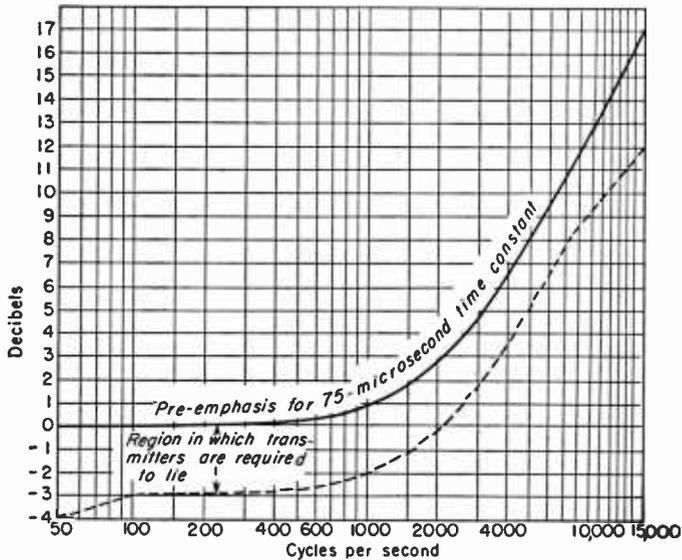


FIG. 29. Preemphasis of upper audio frequencies in f-m broadcasting and in sound channel of television.

tode amplifier. At the lowest frequencies the impedance of the inductance will be negligible, so that the plate voltage is merely IR . At the high frequencies the reactance of the coil will greatly exceed the value of the fixed resistor, so that the plate voltage is almost equal to IX_L . The plate voltage, therefore, will gradually increase for a given current as higher audio frequencies are applied. The 75- μsec provision merely means that the ratio L/R must be 75×10^{-6} ; *i.e.*, there must be 75 μh per ohm of resistance or the equivalent relationship of 13.3 ohms of resistance per millihenry of inductance.

The extent to which the upper frequencies are reinforced by a circuit of 75 μsec time constant is shown by the solid curve in Fig. 29. Since some tolerance is necessary with actual equipment, the FCC requires each station not to have greater preemphasis than this and allows them to have less preemphasis down to the dashed line in the figure. The following formula can be used for computing preemphasis at any frequency:

$$(\text{db})_p = 20 \log_{10} \sqrt{1 + 0.222 \times 10^{-6} f^2}$$

where f is in cycles. This formula, like the solid curve in Fig. 29, gives the preemphasis at any frequency in comparison with the lowest audio frequencies.

The same amount of preemphasis is used in the sound channels of television broadcast stations.

At the receiver it is necessary to introduce a h-f loss, *i.e.*, a deemphasis, to restore the audio frequencies to their proper balance and to realize the advantage in noise reduction. It is usually the practice in receiver design to observe the amount of loss for the upper frequencies present due to the natural characteristics of various elements and then to add a shunt capacitor at one point of such value as to increase the loss to the necessary amount. As a matter of theoretical interest, it may be noticed that, should all the necessary deemphasis loss be inserted at one point, an *RC* combination of 75 μsec would be required, such as furnished by 50,000 ohms in parallel with 1,500 $\mu\mu\text{f}$ or by 25,000 ohms in parallel with 0.003 $\mu\mu\text{f}$.

134. RMA Standards for F-m Receivers. The RMA has standardized upon a balanced 300-ohm transmission line for the input connection of f-m broadcast receivers. A parallel-wire construction with a continuous low-loss plastic separator is available for this service.

The standard i.f. is 10.7 Mc. This value slightly exceeds half of the 20-Mc total width of the f-m band, which is 88 to 108 Mc, so that no f-m station will lie on the image when another is being received. The possibility, of course, still remains that image interference may be encountered from other services outside the 88–108 Mc band.

Measurement procedures for f-m receivers are considerably different from those for a-m sets (see Sec. 22).

135. Application of Negative Feedback to Oscillator of F-m Set. If the output audio voltage delivered by an f-m receiver is used to produce f.m. of the local heterodyne oscillator of the receiver with such a direction as to reduce the frequency difference between the oscillator and the incoming signal, a form of negative feedback is obtained¹ in which the frequency swing of the i-f signal is reduced. In this way distortion originating in the receiver can be greatly reduced, so that the method is of special interest in connection with multiplex f-m systems, such as used in point-to-point practice.²

136. F-m Broadcast Receivers. Although numerous techniques are available for the design of receivers for f-m broadcasting, experience has shown that acceptable results at the lowest cost are obtained with a rather standardized group of features. These include the superheterodyne circuit tuned by a capacitor gang with the oscillator frequency higher than the signal. An r-f stage is generally used, and the more selective sets have a three-gang capacitor for tuning the antenna circuit, the interstage coupling, and the oscillator. A less selective design is also widely made in which a two-gang capacitor tunes the interstage coupling and the oscillator; *i.e.*, the antenna circuit has a fixed broad tuning for the entire band from 88 to 108 Mc. The use of the r-f stage reduces oscillator radiation. The two-gang design radiates less than the three-gang because the common coupling in the rotor shaft of the capacitor does not, in the two-gang case, afford a route for oscillator energy to reach the antenna circuit.

137. Capacitive vs. Inductive Tuning. A minor disadvantage of capacitive station tuning is that the common coupling in the capacitor rotor shaft in a three-gang design limits the gain of the r-f stage to about 20 db. The capacitive design is, however, generally used for other reasons.

One inductive design³ with continuous tuning over the television and f-m bands from 44 to 216 Mc has three tuning coils on a common coaxial shaft of insulating material. Contacts are provided which short-circuit part of each coil as the shaft

¹ Chaffee, J. G., The Application of Negative Feedback to Frequency-modulation Systems, *Proc. IRE*, May, 1939, pp. 317–331.

² Clavier, A. G., and G. Phelizon, Paris-Montmorency 3000-Mc F-m Link, *Elec. Commun.*, June, 1947, pp. 159–169.

³ Ware, Paul, Inductive Tuning System for FM-television Receivers, *Proc. Radio Club of America* May 1946, pp. 9–15. Melvin, M. F., Variable Inductance Tuning for TV Receivers, *Communications*, April, 1947, pp. 48–49, 63.

is rotated through 10 turns to cover the frequency range. Two halves of a 6J6 double triode connected in parallel act as a grounded-grid r-f amplifier, a 6AK5 pentode is used as mixer, and half of another 6J6 is used as oscillator. Two of the tuned circuits are used as a band-pass coupling between the amplifier and the converter, while the third tunes the oscillator. This design gives a total gain of 19 to 23 db over its frequency range with a band width of 5 to 7 Mc and an image ratio of 32 to 45 db, the i.f. being 21.9 Mc.

138. Design Points on F-m Broadcast Receivers. The 100-Mc f-m band is high enough for the grid losses of tubes to be important. Measurements¹ at 100 Mc under typical operating conditions rate the 6AK5, 6AG5, 9003, and 6SH7 in order of decreasing merit as a means of producing r-f gain from antenna to the tube plate. Tubes suitable for use as r-f amplifiers include the 6AK5, 6AG5, 6SG7, and 6BA6 pentodes, and the 12AT7 double triode. For converter, suitable types are the 7F8 and 12AT7 double triodes, the 6AK5 pentode, and the 6SB7Y and 6BE6 heptodes. As separate oscillator, if one is used, the 6C4 triode and the 7F8 are suitable.

The i-f and detector design may consist of two i-f stages, followed by a grid-current limiter and a discriminator. Such a limiter, however, does not reject a.m. of fairly high frequency, such as the 400- or 600-kc undulation due to interference from the second or third channel away in frequency allocation; in fact, these limiters increase such interference. This drawback is avoided and one tube saved by using only one i-f stage, followed by a driver stage and a ratio detector.

139. Tuning Indicators. High-grade f-m receivers should be provided with a tuning indicator. Medium-price sets can be tuned fairly well by the user without a tuning indicator if (1) the selectivity curve falls off about 6 db at ± 75 kc, (2) no limiters are used, (3) a ratio detector is used, and (4) only a limited amount of a-v-c action is provided. Various circuit arrangements² for f-m tuning indicators are available.

A simple tuning-indicator circuit for receivers having limiters is one which develops a control voltage from the grid current of a limiter stage. An example is a tuner³ which has a 6E5 electron-ray tube operated from the grid-return circuit of a 6SJ7 used as first limiter stage.

A tuning meter can be used,⁴ where a 50- μ amp zero-center meter is connected in series with the 470,000-ohm resistor between cathodes of the 6H6 phase-type discriminator.

The 6AL7GT electron-ray tube with three deflecting plates is specially suitable for f-m receivers and is discussed in Sec. 92.

140. F-m Detectors. The phase-variation, discriminator, or "coupled-circuit" type of f-m detector is similar to the discriminator of an a-f-c system described in Sec. 97. While the carrier frequency is not being deviated, this detector rejects a.m. by virtue of the equal and opposite rectified voltages in its output circuit. However, when the frequency is at some deviated value, one rectified voltage exceeds the other, and it can be seen that an increase of level due to a.m. will increase both rectified voltages and therefore their difference. For example, if the voltages were -5 and $+6$, with no a.m., and were doubled by the a.m. to -10 and $+12$, the differential output would be increased from 1 to 2 volts.

141. Ratio Detectors. Detectors⁵ having the property of being insensitive to a.m., while the carrier is deviated as well as when it is not, have been developed. Of special interest is the ratio detector,⁶ which has this property and obtains its name because of

¹ From *Application Note 118*, Apr. 15, 1947, issued by Tube Department, RCA, Harrison, N.J.

² Rodgers, John A., Tuning Indicators and Circuits for Frequency-modulation Receivers, *Proc. IRE*, March, 1943, pp. 89-93.

³ Spindel, F. A., Data on the Browning FM-AM Tuner, *FM & Television*, February, 1947, pp. 34-38.

⁴ Cosman, B. J., and A. W. Richardson, Harvey Double Superheterodyne FM Receiver, *FM & Television*, June, 1947, pp. 21-24, 51-52.

⁵ Beers, G. L., A Frequency-dividing Locked-in Oscillator Frequency-modulation Receiver, *Proc. IRE*, December, 1944, pp. 730-737. Bradley, W. E., Single-stage F-M Detector, *Electronics*, October, 1946, pp. 88-91. Loughlin, Bernard D., Performance Characteristics of FM Detector Systems, *Tele-Tech*, January, 1948, pp. 30-34.

⁶ Seeley, S. W., and Jack Avins, The Ratio Detector, *RCA Rev.*, June, 1947, pp. 201-236. This paper

its apparent sensitiveness to the ratio of the two voltages developed from the applied signal, rather than to their difference. Since the ratio between the two voltages is the same at any level, an ideal ratio detector does not respond to a.m. at any frequency deviation.

There are two general types of ratio detectors, the essential parts of which are shown in Fig. 30. The phase-shift type is the more widely used. Fundamentally, the two types differ only in the use of a different i-f coil structure between the last i-f stage and the diodes. In common with discriminators, ratio detectors make use of phase shift or of side tuning to develop the necessary frequency-sensitive i-f voltages. The tuning of the side circuits in the side-tuned type is shown in Fig. 30B by the notation f_a and f_b to indicate the resonant frequencies of the two side-tuned circuits, one above and one below the center value of the i.f., f_i .

The ratio detector differs from the discriminator in the use of a large capacitor C which maintains a constant voltage during all a-f variations. The time constant of C , R_1 , and R_2 has a typical value of $\frac{1}{2}$ sec. For example, C may be $8 \mu\text{f}$ and R_1 and R_2 may be 6,800 ohms each.

Considerable care is necessary in the design of the transformers and detail circuits of ratio detectors to secure good results. Accurate balance and other refinements are necessary.

A reasonable sensitivity requirement for a driver i-f stage and ratio detector operating at an i.f. of 10.7 Mc is that a signal level of 0.1 volt rms on the driver grid with a 22.5-kc (30 per cent) deviation should produce an audio output of about 0.2 volt rms.

The characteristics of the f-m detector in a receiver have an important effect on the apparent tuning action. For example, if the

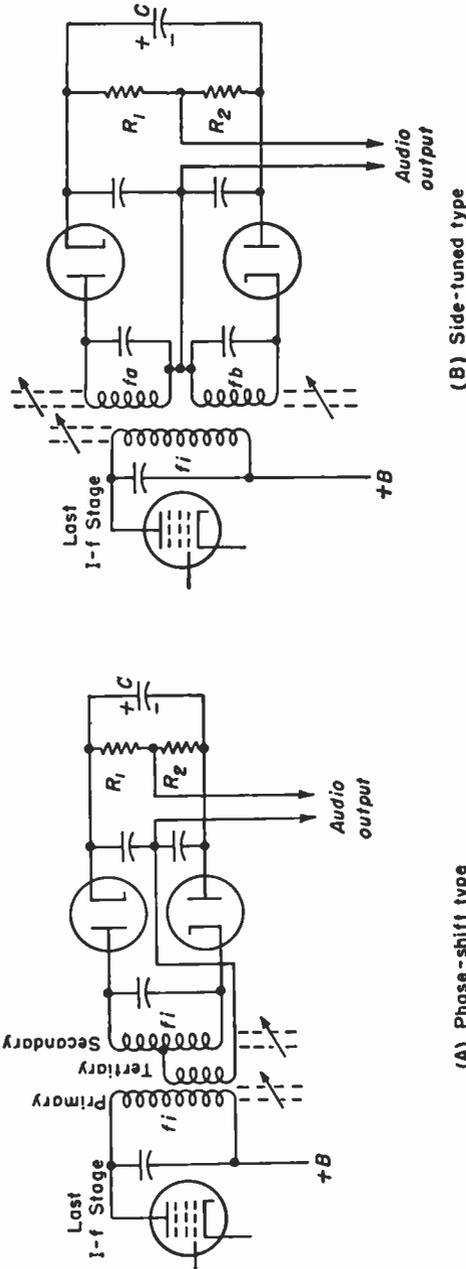


Fig. 30. Essential parts of the two general types of ratio detectors.

is also partly given under title Ratio Detector for FM Signals, *Tele-Tech*, July, 1947, pp. 46-49. Hayes. Albert T., Jr., Improved Type of Ratio Detector, *Tele-Tech*, November, 1947, pp. 41, 94, 96.

over-all i-f band width at 6 db down is between 150 and 200 kc, the tuning action will be good if the separation of the detector peaks is 300 to 400 kc (± 150 to ± 200); otherwise a noticeable false response may occur on each side of the correct tuning point. The separation of the detector peaks can be observed by using a battery and potentiometer to hold across C a voltage corresponding to the signal strength of interest and then applying in succession various unmodulated i-f signals over a frequency range extending beyond the peaks in each direction.

Theoretical and experimental work¹ confirms that the rejection of a.m. by a ratio detector is produced in a different way from ordinary limiting, such as accomplished

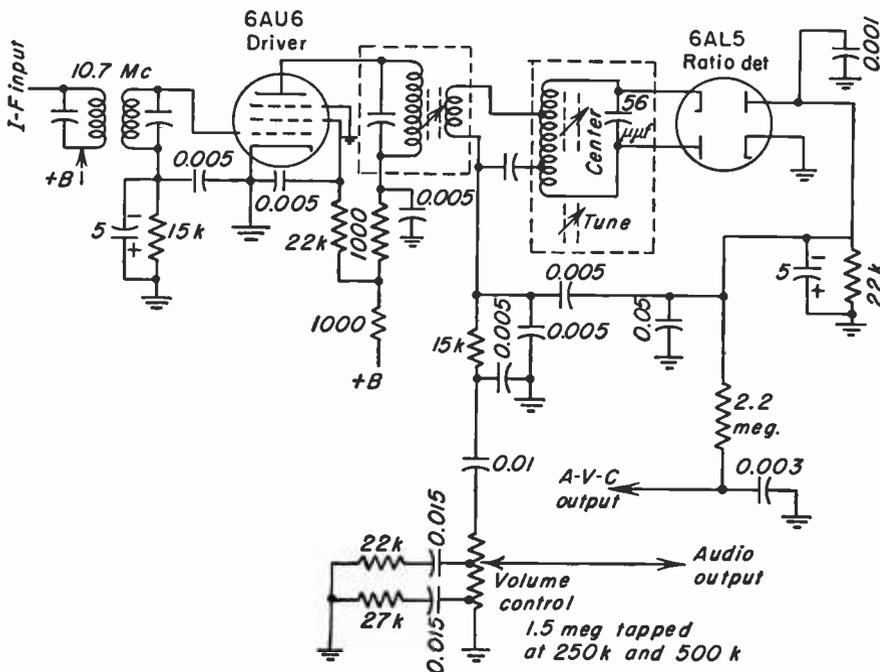


FIG. 31. Ratio detector used in RCA Model 68R1.

by conventional grid-circuit limiters which remove a.m. from the carrier wave. An oscilloscope may show some a.m. on the carrier wave to be removed by a ratio detector; however, this action is incidental and unrelated to the a-m rejection properties. Ratio detectors can be designed to achieve complete a-m rejection when the apparent limiting action is very small. While a diode dynamic limiter (see Sec. 142) has best limiting action with the highest diode conductance, practical ratio-detector circuits usually add resistance in series with each diode to obtain optimum a-m rejection.

Germanium crystals can be used in place of vacuum diodes in ratio detectors, but may require added series resistance to reduce the forward conductance. They also have the limitation that the backward conductance may increase at high signal levels and thereby impair the detector performance.

The grid-to-plate voltage gain of the driver can be substantial. A reasonable maximum figure can be obtained by subtracting 8 or 10 db from the result given by the following expression:

$$20 \log_{10} \left(2 \times \frac{\text{reactance of grid-plate capacitance}}{\text{resistive component of grid impedance}} \right)$$

¹ Loughlin, B. D., "The Theory of Amplitude Modulation Rejection in the Ratio Detector," paper at IRE Convention, New York, March, 1950.

This formula gives the maximum gain, measured at resonance, for a ratio-detector driver that will just start to oscillate when it is detuned during alignment to the point where maximum feedback occurs. For example, the maximum safe gain at 10.7 Mc for a 6BA6 tube ($C_{gp} = 0.0035 \mu\mu\text{f}$) in a circuit having an effective grid resistance of 36,800 ohms (Q of 72, C of $30 \mu\mu\text{f}$) would be 46.8 db. The driver gain in this case

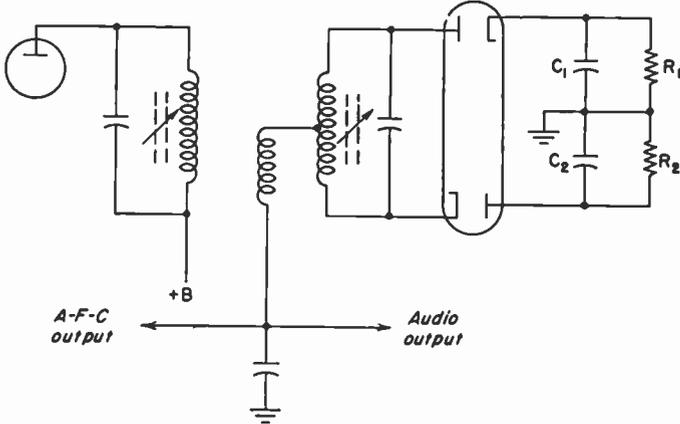


FIG. 32. Ratio detector delivering audio and a-f-c outputs in Hallicrafter chassis.

should not, therefore, be allowed to exceed 37 to 39 db. Note that the value of the resistive component of the resonant impedance in the driver external grid circuit when the primary is well detuned is to be used as the denominator in the formula.

The circuit of the ratio detector in the RCA Model 68R1 is given¹ in Fig. 31. The RC circuit has a time constant of 0.11 sec. Ratio detectors with a single three-winding transformer between the driver and the diodes are shown in Figs. 32 and 33.² The

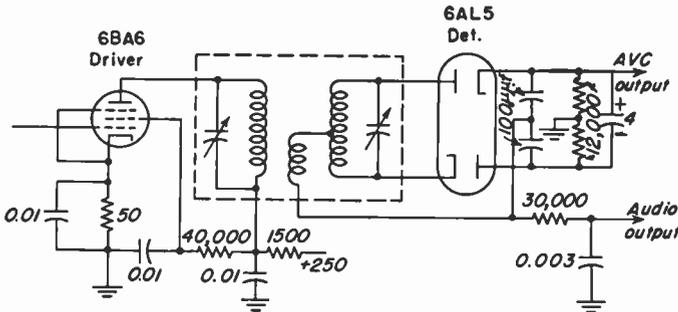


FIG. 33. Ratio detector described by Hobbs.

Hallicrafters design furnishes a-f-c voltage³ as well as audio output. The use of push-button tuning in the 100-Mc f-m band is made practical in this way. The time constants R_1C_1 and R_2C_2 are long compared with the period of the lowest a.f. to be handled, and R_1 and R_2 are large compared with the forward internal resistance of each diode.

The ratio detector shown in Fig. 33 is the basis for some practical discussion by

¹ F-M/A-M Ratio-detector Table Model Receiver. *Service*, May, 1947, cover and pp. 13 and 50.
² An additional ratio-detector circuit used in manufactured equipment is given in Data on the F-M Pilotuner, *FM & Television*, September, 1947, pp. 37, 40.
³ Hershey, Lloyd M., Design of Tuners for AM and FM, *Tele-Tech*, August, 1947, pp. 58-59, 95.

Hobbs in the literature.¹ He points out that careful design and factory production are necessary to obtain low sensitivity to a.m. and low audio distortion. For these purposes the two diode circuits must be carefully balanced. The use of bifilar secondary windings of high Q is beneficial in this regard. Also, the i-f amplifier must not be too selective; otherwise it will produce excessive a.m. of which the downward modulation may be too severe to be rejected. It is concluded that, when well designed and accurately produced, a ratio-detector set will be insensitive to a.m. at lower signal levels than will a set having a limiter and phase-variation discriminator.

142. Dynamic Limiters. The grid-bias type of limiter cannot be given a short enough time constant to follow the envelope and eliminate a.m. of high frequency, such as the 400 kc encountered in second channel interference. This difficulty can be avoided by the use of a "dynamic limiter" which utilizes one or two diodes that respond instantly to amplitude variations and limit them. Like a ratio detector, it has an RC circuit which holds its characteristics constant with regard to a-f variation but permits slow change to adjust to the signal level. The dynamic characteristic

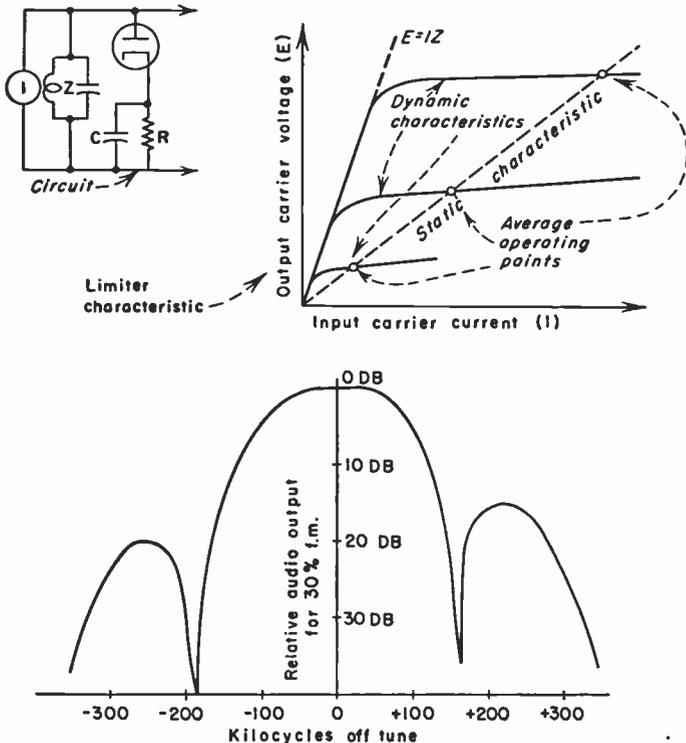


FIG. 34. Dynamic limiter for f-m receiver.

curves (Fig. 34) are therefore almost flat, while the static characteristic is a linear output-input relation. Used with a phase-variation discriminator (which is not a ratio detector), the dynamic limiter gives the receiver the tuning action shown in Fig. 34. The side responses are fairly weak, being about 17 db down. There is relatively little interstation noise, a property similar to that of a ratio detector.

Practical further utilization of the dynamic limiter can be made by the substitution of a germanium crystal in place of the vacuum diode. By the use of two diodes or

¹ Aram, Nathan W., Lloyd M. Hershey, and Marvin Hobbs, F-m Reception Problems and Their Solution, *Electronics*, September, 1947, pp. 108-113. The treatment of f-m detectors here is by Hobbs.

crystals with opposite polarity, both halves of the modulation envelope can be controlled. Also the dynamic limiter can be used with a ratio detector with very good performance and a relaxation of the requirements on each. It is worth noting that the use of a dynamic limiter with a ratio detector affords the best performance now obtainable in an f-m receiver.¹

HIGH-FIDELITY RECEIVERS.

Most radio receivers produce a quality of sound which leaves a good deal to be desired in comparison with the original music or speech. The defects include restriction of the frequency and amplitude ranges and the introduction of hum, noise, and extraneous harmonic and intermodulation frequencies. Receivers which substantially surpass the average set in the quality of reproduction are called "high-fidelity receivers."²

143. Characteristics of Ear and of Music, Speech, and Sound Effects. The characteristics of the ear³ for persons twenty years of age include a frequency range from 20 to 16,000 cycles with the greatest sensitivity over the portion from 1,000 to 5,000 cycles. In this central range an average twenty-year-old listener can hear sounds over the large intensity range of 105 db at 1,000 cycles, 110 db at 2,500 cycles, and 105 db at 5,000 cycles. For higher or lower frequencies, the intensity range of the ear greatly declines as this central frequency region is departed from. This intensity range is, however, not fully utilizable in listening to the average receiver operating with a loud-speaker because of interfering room noise which is always present and masks faint sounds that would otherwise be heard. The sensitivity of the ear for the higher frequencies declines with increasing age, especially for men so that at age fifty-five the sensitivity at 7,040 cycles is -24 db for men and -19.7 db for women.⁴

Although frequencies as low as 20 cycles and components as high as high as 15,000 cycles exist in music, an analysis by Fletcher⁵ leads to the conclusion that under practical circumstances substantially complete fidelity for symphonic music is obtained with a volume range of 65 db and a frequency range from 60 to 8,000 cycles. For speech the corresponding figures are 40 db and 100 to 7,000 cycles.

The single-channel character of the normal radio system produces what may be called a "monaural" type of operation, so that the binaural or "stereoscopic" feature made possible by the two ears of the listener is not used. This is a standard limitation which has been removed in special demonstrations.⁶

144. Preferences of Listeners. There is a difference of opinion as to the reason why a typical listener removes the higher pitches of music with his tone control. A careful analysis of various data and procedures in this connection, published by LeBel,⁷ should be studied by all who are concerned about the matter.

A direct acoustic test of listener preference for the upper frequencies was made by Olson,⁸ the apparatus including no electrical equipment. Original music and speech

¹ For other references to f-m receivers and systems, see Hazeltine FreModyne FM Circuit, *Tele-Tech*, December, 1947, pp. 41, 85-86; FreModyne F-m Receivers, *Electronics*, January, 1948, pp. 83-87; A. R. Boone, The Army's Radio Relay Equipment, *Radio News*, January, 1946, pp. 25-28, 151-155 (this describes AN/TRC-1, AN/TRC-8, and other models); G. G. Gerlach, A Microwave Radio Communication System, *RCA Rev.*, December, 1946, pp. 576-600; W. F. Goetter, Frequency-modulation Transmitter and Receiver for Studio-to-transmitter Relay System, *Proc. IRE*, November, 1943, pp. 600-606; D. E. Noble, A State-wide F-M Police System, *Electronics*, November, 1940, pp. 18-22, and December, 1940, pp. 28-31, 66 (The receiver is described in the second installment).

² Music in the Home, *Fortune*, October, 1946, pp. 155-163, 190, 193, 195. Hanson, O. B., Comments on High Fidelity, *Electronics*, August, 1944, pp. 130-131, 385-391.

³ Fletcher, Harvey, Hearing, the Determining Factor for High-fidelity Transmission, *Proc. IRE*, June, 1942, pp. 266-277.

⁴ Steinberg, J. C., H. C. Montgomery, and M. B. Gardner, Results of World's Fair Hearing Tests, *Bell System Tech. J.*, October, 1940, pp. 533-562.

⁵ *Loc. cit.*

⁶ *J. Soc. Motion Picture Engrs.*, October, 1941, pp. 331-426.

⁷ LeBel, C. J., Psycho-acoustical Aspects of Listener Preference Tests, *Audio Eng.*, August, 1947 pp. 9-12, 44-48.

⁸ Olson, H. F., Frequency Range Preference for Speech and Music, *Electronics*, August, 1947, pp. 80-81; also given under title, Report on Dr. H. F. Olson's Listener Preference Tests, *Audio Eng.*, June, 1947, pp. 27, 43-44.

with a full frequency range were compared with the same sound except that frequencies above 4,000 cycles were removed. An acoustic filter, made up of large panels that could be rotated in and out of position, was used with an opaque sound-transparent screen to hide it from the audience. The room had dimensions 24 ft long, by 20 ft wide, by $9\frac{1}{2}$ ft high. The musical tests were made with a six-piece orchestra in one corner and the listeners in the opposite corner. For speech tests the orchestra was replaced by one person. Over a thousand listeners participated in the tests and were told that there was a source of music or speech behind the curtain and that the program would be rendered under two different conditions, indicated by the display of letters *A* or *B* on the indicator, and that these conditions would be changed every 15 sec. The listeners were asked to vote on which condition they considered preferable. About two-thirds of the observers preferred the wide range, the figures being about the same for dance music, semiclassical music, and speech. At a meeting of the Acoustical Society of America, J. P. Maxfield, Bell Telephone Laboratories, reported that with electronic systems free from h-f transients and cross modulation, he had observed listeners to prefer wide-band reproduction. This, he stated, was true for both single-channel and stereophonic systems.

Extension of the upper audio range greatly increases the requirements for low noise and low introduction of harmonics and intermodulation frequencies. If these faults are present, the upper register will detract from, rather than improve, the over-all quality, and the average listener will unhesitatingly turn down the tone control.

145. Fidelity of A-m and F-m Receivers. The 10-kc spacing of stations in the standard broadcast band places a definite upper limit on the reproduction of the higher audio frequencies. This is 4,000 or 5,000 cycles for distant reception and may be substantially higher for strong local signals.

When one end of the audio range of the receiver is restricted, the most natural results are obtained by proportional restriction at the other end of the frequency range; *i.e.*, the l-f range should be extended downward in the same relative proportion as the upward extension of the h-f range. A common working rule is that the product of the lowest and highest frequency limits should be approximately 500,000. The following table gives typical ranges¹ and the corresponding frequency product for four types of receivers:

Type	Lower frequency limit	Upper frequency limit	Frequency product
Small table set.....	190	2,700	514,000
Large table set.....	140	3,500	490,000
Typical console.....	100	4,500	450,000
High-fidelity console.....	80	6,800	528,000

If the frequency range goes above 8,000 cycles, an a-m receiver is likely to bring in "monkey chatter," produced by undesired side frequencies beating with the desired carrier or by an undesired carrier beating with the desired side frequencies. It is only with very strong local signals that high-fidelity reception above 8,000 cycles can be enjoyed. In this case the next hurdle is the 10-kc whistle between carriers on adjacent channels; this can be removed with a band-elimination filter.

To obtain a receiver that will produce fairly high-fidelity output on local stations and still be useful for the reception of distant signals, variable selectivity should be provided.

An a-m receiver intended for high-fidelity reception should have special attention devoted to the last i-f amplifier to avoid overloading at this point. It may be necessary to reduce or remove the a-v-c bias on this stage. The audio signal is, of course, the envelope of the r-f wave, so that overloading, which would distort the envelope, must be avoided.

¹ Hanson, *loc. cit.*

Another refinement is approximately equal audio and d-c loads for the diode detector. Failure to provide this equality will cause clipping of the output during the inward modulation (also called the "downward" modulation), so that distortion in the reproduced audio wave will occur, especially for modulation exceeding about 50 per cent.

As instances of high-fidelity receivers for the standard broadcast band, reference may be made to the set produced before the Second World War by E. H. Scott¹ of Chicago and to the one designed by Lincoln Walsh² of Elizabeth, N.J.

The f-m transmission standards permit the reception of high-fidelity f-m programs with reasonable ease. The design must include provisions for a-m rejection to remove noise and to remove distortion which otherwise would result from a.m. introduced by the rounding of the tops of the selectivity curves of the i-f transformers. Beyond this, the requirements for high fidelity involve only the problems of audio amplification and the loud-speaker.

146. General Requirements for High-fidelity Set. The chief general requirements to be satisfied are as follows:

1. The set must be practically free of hum.
2. The reproduction must be practically free of harmonics and intermodulation frequencies at all volume levels.
3. The electrical characteristics must be practically uniform over a wide audio range.
4. A good loud-speaker must be provided.

Interesting general treatments of these requirements³ and of the studio and motion-picture techniques⁴ may be consulted in the literature.

A critical portion of an audio amplifier from the fidelity standpoint is the output stage because of the high power level. This stage is normally operated push pull to remove the even harmonics and to remove the d-c flux in the output transformer.⁵ There is a difference of opinion as to the best general type of tube for the output stage. Some favor triodes,⁶ owing to the absence of high-order distortion. Others prefer beam tetrodes,⁷ with the addition of negative feedback. If triodes are used, care is needed to avoid overloading of the driver stage that furnishes the input for the final stage.

Whatever tubes are used in the last stage, negative feedback is available as a means of reducing distortion. By this powerful tool quality can be improved with little circuit complication, the only important drawback being the reduction in the gain of the amplifier which can be restored by higher gain couplings or an additional stage.⁸

147. Intermodulation Distortion. Output-input characteristic can be expressed as a mathematical power series in terms of the first, second, third power, etc., giving the instantaneous output as a function of the instantaneous input signal. The second-

¹ Scott, E. H., *Scott High-fidelity Receivers*, Proc. IRE, June, 1941, pp. 295-299.

² High-quality Radio Reception, *Electronics*, July, 1939, pp. 20-23. Edinger, J. R., High-quality Audio Amplifier with Automatic Bias Control, *Audio Eng.*, June, 1947, pp. 7-9, 41.

³ Toth, Emerick, High-fidelity Reproduction of Music, *Electronics*, June, 1947, pp. 108-113. Minter, Jerry, Audio Distortion in Radio Reception, *Proc. Radio Club of America*, January, 1946, pp. 1-5; also available as Chap. 6 of "Standard F-M Handbook," FM Company, Great Barrington, Mass. Hanson, loc. cit.

⁴ Gurin, H. M., and G. M. Nixon, A Review of Criteria for Broadcast Studio Design, *J. Acoust. Soc. Am.*, May, 1947, pp. 404-411. Maxfield, J. P., Liveness in Broadcasting, *West. Elec. Oscillator*, January, 1947, pp. 3-7, 26-27. (In Fig. 4 of this paper the largest three ordinates should be 1.0, 2, and 3 rather than 1.0, 1.2, and 1.3). Content, Edward J., The Acoustical Design of FM Studios, *Tele-Tech*, April, 1947, pp. 30-34. "Motion Picture Sound Engineering," Research Council of Academy of Motion Picture Arts and Sciences, Van Nostrand, 1938.

⁵ Kraus, H. L., Class-A Push-pull Amplifier Theory, *Proc. IRE*, January, 1948, pp. 50-52.

⁶ See Walsh reference mentioned in Sec. 145 and Toth reference mentioned in Sec. 146. Beggs, G. E., Bridging Amplifier for F-M Monitoring, *Electronics*, January, 1946, pp. 152-155.

⁷ Hilliard, John K., High-fidelity FM Reproduction, Chap. 7, pp. 99-103, of "Standard FM Handbook," FM Company, Great Barrington, Mass., 1946. Singer, Kurt, A High-quality Recording Power Amplifier, *J. Soc. Motion Picture Engrs.*, June, 1947, pp. 560-568. Goodell, John D., Amplifiers for Music Reproduction, *Radio News*, April, 1947, pp. 6-8, 27.

⁸ Audio Design Notes: Negative Feedback Circuits, *Audio Eng.*, May, 1947, pp. 35, 52. Bode, H. W., "Network Analysis and Feedback Amplifier Design," Van Nostrand, 1945. This book gives a thorough analysis of feedback amplifiers and networks.

power term produces the second harmonic and the sum and difference of the various applied frequencies; the third-power term produces the third harmonic and with two input frequencies also produces outputs of $2f_1 \pm f_2$ and $2f_2 \pm f_1$. Counting the two harmonics, the result of third-power distortion is, therefore, six undesired frequencies. Similarly, in the case of fifth-power distortion, two input frequencies produce their fifth harmonics and eight intermodulation products. These frequencies resulting from intermodulation are generally inharmonic or discordant and also likely to fall in the middle or upper portion of the audio range, both of which conditions make them quite objectionable. Intermodulation can be quickly measured with suitable equipment.¹

148. Audio Transformers for High-fidelity Sets. The output transformer coupling the last tube to the speaker is one of the worst offenders from the standpoint of fidelity. For adequate linearity to avoid intermodulation, much more iron is needed in the magnetic circuit. Both the quality and the quantity of iron used must, of course, be considered.² Experience has shown that good frequency range and low amplitude distortion are difficult to obtain in transformers required to have several connections for matching to various load impedances. These should in general be avoided in high-fidelity designs.

149. Dividing Networks. Conversion of the a-f signals into sound is facilitated by dividing the frequency range into a lower and an upper part with a separate special speaker for each portion. In this case the h-f speaker is often called a "tweeter," and the l-f speaker a "woofer." A filter network³ is generally provided to route the proper frequencies to each speaker.

Numerous circuit arrangements are available for dividing networks, such as the mere use of series capacitors in the tweeter circuit. More elaborate networks include two filters which may be designed for series or parallel input connection. For the more elaborate designs a choice of the dividing frequency is required; *i.e.*, the frequency at which an equal electrical output is delivered to the two speakers. Above this frequency, the woofer receives decreasing amounts of signal, and below this frequency the tweeter receives decreasing amounts. The dividing frequency varies considerably between different designs and may be anywhere from 250 to 2,000 cycles. The lower values of dividing frequency are generally used in motion-picture practice and the higher values for studio monitoring and high-fidelity home reproduction. The networks used in motion-picture practice generally introduce an attenuation of 12 db per octave as the frequency departs from the dividing frequency. Some experimenters favor locating the dividing network between the driver and the last stage, so as to operate at a lower power level. In this case separate output stages are necessary for the two frequency ranges.

150. High-fidelity Loud-speakers. Acoustics and loud-speakers are treated in Chap. 16; only a few points are discussed here.⁴ The tweeter is often mounted in the

¹ Hilliard, John K., Intermodulation Testing, *Electronics*, July, 1946, pp. 123-127. Pickering, Norman C., Measuring Audio Intermodulation, *Electronic Inds.*, June, 1946, pp. 56-58, 124-125. Daniel, George, Instrument for Intermodulation Measurements, *Electronics*, March, 1948, pp. 134, 136, 150.

² Harrison, E. B., High-quality Communication and Power Transformers, *J. Soc. Motion Picture Engrs.*, September, 1944, pp. 155-167. Partridge, Norman, Harmonic Distortion in Audio-frequency Transformers, *Wireless Engr.*, September, 1942, pp. 394-400; October, 1942, pp. 451-456; November, 1942, pp. 503-507.

³ "Motion Picture Sound Engineering," Research Council of the Academy of Motion Picture Arts and Sciences, Van Nostrand, 1938; see index for various references on this subject. Schuler, E. R., Design of Loudspeaker Dividing Networks, *Electronics*, February, 1948, pp. 124-125. McProud, C. G., Design and Construction of Practical Dividing Networks, *Audio Eng.*, June, 1947, pp. 15-17, 45-46. Klipsch, Paul W., Woofer-Tweeter Crossover Network, *Electronics*, November, 1945, pp. 144-145. Design Data for Dividing Networks, *Aerovox Research Worker*, October, 1945, published by Aerovox Corp.

⁴ Among references of interest are the following: Goodell, John D., Loudspeakers for Extended Range Applications, *Radio News*, May, 1947, pp. 3-5, 30-31; June, 1947, pp. 6-8, 29. General treatment is given plus information on Altec-Lansing, Klipschorn, and RCA speakers. Klipsch, Paul W., A Note on Acoustic Horns, *Proc. IRE*, July, 1945, pp. 447-449; also The Klipsch Sound Reproducer, *FM & Television*, September, 1947, pp. 25-27, 46, 48. Jensen Technical Monograph series: (1) Loud Speaker Frequency-response Measurements; (2) Impedance Matching and Power Distribution; (3) Frequency Range and Power Considerations in Music Reproduction; (4) The Effective Reproduction of Speech; and (5) Horn Type Loud Speakers. These are available from Jensen Radio Mfg. Co., Chicago.

opening of the woofer, a coaxial arrangement with the advantages of compactness and identity of source location for the two ranges of frequencies.

Diffusing vanes for the tweeter are desirable to correct the beamlike character of the h-f radiation. An example of a coaxial speaker with such vanes is the Altec-Lansing Duplex.¹ This speaker is used with a dividing network having a crossover frequency of 1,200 cycles. With a baffle of 6 cu ft volume, good response is obtained down to 60 cycles and with 9 cu ft down to 40 cycles. The inner surface of the baffle must be covered with sound-absorbent material to prevent reflections that would give a "hang-over" or echo effect. The vanes provide for the h-f radiation, giving a distribution of 60 deg horizontally by 40 deg vertically.

Single-unit speakers for high-fidelity use have been designed.²

151. Variable Selectivity. Ability to vary the selectivity of a receiver is useful where wide variations in receiving conditions are encountered. In broadcast reception, variable selectivity permits the use of a wide carrier-frequency pass band for the high-fidelity reception of local programs, while upon contraction of the selectivity distant stations can be received free of cross talk. In amateur practice, very sharp selectivity is often required to receive phone and telegraph messages in the presence of intense interference. Amateur receivers generally include variable selectivity with an adjustable quartz-crystal filter. By this means band widths as small as 100 cycles can be obtained.

Variable selectivity is easily obtained in the superheterodyne receiver by a change of the characteristics of the i-f amplifier. Available methods are (1) variation of the coupling between the windings of one or more of the i-f transformers so as to widen the selectivity through over-optimum coupling;³ (2) introduction of damping, *i.e.*, lowering the *Q* of one or more of the coupling systems (this has the disadvantage of reducing the selectivity against interference having a frequency some distance from the carrier frequency); (3) detuning the i-f tuned circuits, so as to convert the amplifier to staggering; and (4) use of an auxiliary feedback tube which can be made to cut down the peak and build up the skirts of the selectivity curve.

Some methods produce symmetrical variation of the selectivity while others produce unsymmetrical variation, the distinction being whether the selectivity curve for the various adjustments is the same or different on the two sides of resonance. The following table⁴ compares the characteristics of the two types of operation:

Characteristic	Symmetrical type	Unsymmetrical type
Complication of operation.....	Only one variable, the amount of expansion	Two variables, <i>i.e.</i> , the response on the two sides
Fidelity.....	Not critical	May be critical because of carrier being near edge of band
Selectivity.....	Better when comparable interference is encountered on both sides of the carrier	Better when major interference is on only one side
Noise.....	Minimum	About 3 db above minimum
General usefulness.....	Superior from fidelity standpoint	Superior from selectivity standpoint

A laboratory receiver with a continuous manual variation of selectivity was constructed at Hazeltine and is described in the paper⁴ by Wheeler and Johnson. This

¹ Lansing, James B., The Duplex Loudspeaker, *J. Soc. Motion Picture Engrs.*, September, 1944, pp. 168-173.

² Olson, H. F., and J. Preston, Wide Range Loudspeaker Developments, *RCA Rev.*, June, 1946, pp. 155-178. Nickel, Frank, Quality Loudspeakers for Every Use, *West. Elec. Oscillator*, July, 1947, pp. 6-10, 30. Lanier, R. S., What Makes a Good Loudspeaker? *West. Elec. Oscillator*, July, 1947, pp. 11-13, 34-35.

³ Variable selectivity obtained by change of coupling is discussed in K. R. Sturley, "Radio Receiver Design," Wiley, pp. 306-321; and also the treatment of automatic variable selectivity on pp. 332-335.

⁴ Wheeler, H. A., and J. K. Johnson, High Fidelity Receivers with Expanding Selectors, *Proc. IRE*, June, 1935, pp. 594-609.

design uses variation of coupling in two i-f transformers, of equal Q , and one additional interstage coupling of $Q/2$ to obtain an approximately flat-topped selectivity curve. The valley between peaks caused by the overcritical coupling of the adjustable transformers is filled in by the single peak of the $Q/2$ transformer. Broadcast receivers with variable selectivity generally provide two or three band widths. One or two small windings are used in each variable transformer to increase the coupling to the desired overcritical value. When only the regular windings are in use, the transformer has critical or slightly undercritical coupling, and the selectivity is sharpest.

Laboratory receivers have been made in which the selectivity curve automatically expands as stronger signals are tuned in, and contracts as weaker signals are received. In a design described by G. L. Beers¹ triode tubes reduce the Q of various windings of the coupling transformers.

A discussion of automatic selectivity control by H. F. Mayer² compares variation of coupling, variation of Q , and detuning, and leads to the conclusion that alteration of the coupling produces the best results for the additional parts required.

152. Use of Interference for Automatically Adjusting Selectivity. By means of an audio tuned circuit, it is possible to isolate the 10-kc beat between the carriers of adjacent broadcast stations and utilize this to increase the selectivity, thus making the receiver selective to an amount depending upon the strength of the interference.³ This design includes a control tube for each of two i-f stages, the control tube feeding back energy to reduce the peak of the selectivity curve by degeneration and build up the sides by regeneration. For weak signals or strong 10-kc beat, the control tubes are biased toward cutoff. As stronger signals are received, or weaker interference, this negative bias is reduced, thus widening the selectivity.

Another automatic selectivity control⁴ includes two traps tuned to 9 kc above and 9 kc below the carrier frequency. They operate independently to contract the selectivity on the side where interference is experienced. In this way the contraction may be on either side or on both sides.

AUTOMOBILE RECEIVERS

153. Characteristics. The automobile set differs from home sets in that:

1. Use of a small whip antenna necessitates high sensitivity.
2. Limited space in the car requires compact construction.
3. Vibration and wide temperature range necessitate rugged parts and construction.
4. Inductive tuning of the r-f and oscillator circuits is generally used.
5. The 6-volt storage battery drives a vibrator for plate supply.
6. Severe disturbance of the engine ignition requires thorough shielding and by-passing.
7. Metal housing is used, generally single-unit.

An examination of nine 1947 models made by four manufacturers showed the following front-end features:

No. of Models	Type of Circuit
5	Tuned r-f stage
2	Untuned r-f stage
1	Double-tuned preselector without r-f stage
1	Single-tuned antenna circuit without r-f stage

Push-button control was generally provided, the most common number of preset stations being five. Values of i.f. were equally divided between 455 and approximately 262 kc. Capacitive tuning of the i-f transformers was provided in seven models; the remaining two had 455-kc i.f. and employed inductive tuning with a fixed capacitor of 100 μf . Tone control was provided in most of the sets. All models included a

¹ Beers, G. L., Automatic Selectivity Control, *Proc. IRE*, December, 1935, pp. 1425-1440.

² Mayer, H. F., Automatic Selectivity Control, *Electronics*, December, 1936, pp. 32-34.

³ Farrington, John F., Receiver with Automatic Selectivity Control Responsive to Interference, *Proc. IRE*, April, 1939, pp. 239-244.

⁴ Rust, N. M., O. E. Keall, J. F. Ramsay, and K. R. Sturley, Broadcast Receivers; A Review, *J. IEE (London)*, June, 1941, pp. 797; see Sec. 2.6 of the paper.

dynamic speaker with electromagnetic field excitation, and all included a tube-type rectifier for plate supply. Vibrators with high-voltage rectifying contacts have been successfully used in the past, obviating the need for a rectifier tube, but none was included among the nine sets in this group. Likewise permanent-magnet speakers for auto sets are available but were not used in the sets of this group.

Antennas are generally short whip types of low capacitance and limited signal-collecting ability. An article by Cull¹ states that it is necessary to conserve the signal by the use of leadin cable with a capacitance not exceeding 1 μf per in. of length and with polyethylene insulation. Closely woven shielding braid is also needed to prevent the pickup of ignition interference.

The vibrator² in the plate-supply circuit commutates the d.c. from the battery so that flux in the transformer reverses direction in step with the moving reed. The frequency may be 115 cycles as a typical value. The transformer steps up the voltage, which is then rectified to serve as the desired plate supply.

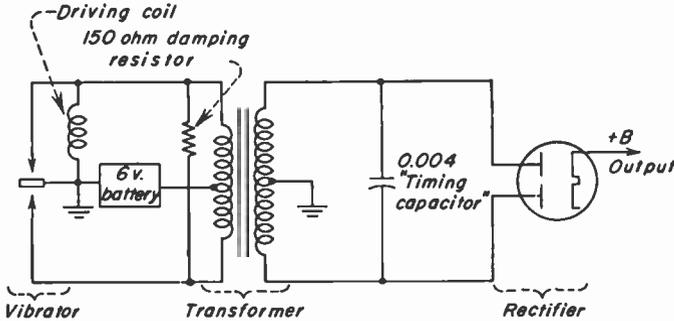


FIG. 35. Typical shunt-coil type of vibrator with stepup transformer and rectifier tube for furnishing plate-supply voltage. The battery can be connected with either polarity.

Noise suppression in automobile receivers³ often requires the by-passing of the generator. Miscellaneous bonding of car parts and also by-passing, such as the dome-light wiring, may be necessary, especially with old cars. The U. S. Rubber Company has announced⁴ an antistatic powder which, when blown into the inner tubes, will reduce interference which may otherwise be produced by the tires. A small amount of the powder is injected into the deflated tube with an air hose and a special container.

V-H-F AND HIGHER BAND RECEIVERS⁵

As receiver designs are made for the v-h-f range and above, various differences in the general conditions which prevail are found. Specially noticeable are (1) the increase of tube input conductance, (2) the smaller amount of atmospheric disturbance and of

¹ Cull, R. R., Auto Antenna Installation, *Radio Maintenance*, September, 1947, pp. 4-7.

² "Mallory Vibrator Data Book," P. R. Mallory & Co., Indianapolis. Devoted to general operation and design. Briefer treatment available in "Mallory Technical Manual," pp. 64-92, 1942.

³ Blair, Don, Auto Radio Noise Elimination, *Radio Maintenance*, December, 1947, pp. 12-13, 38, 42. Jeep cars made for service during the war had extensive bonding, a triple filter, five 10,000-ohm high-tension suppressors, and two generator by-pass capacitors. See Butler, F. E., Suppressing Radio Noise in the Jeep, *Electronics*, December, 1943, pp. 96-99.

⁴ *Electronics*, July, 1947, pp. 208, 210.

⁵ As general references, see the following: Staff of Radio Research Laboratory, Harvard University, "Very High Frequency Techniques," McGraw-Hill, 1947; E. W. Herold and L. Malter, Some Aspects of Radio Reception at Ultra-high-frequency, *Proc. IRE*, August, 1943, pp. 423-438; September 1943, pp. 491-510; October, 1943, pp. 567-582; G. E. Hulstede, J. M. Pettit, H. E. Overacker, K. R. Spangenberg, and R. R. Buss, Very-high-frequency Receivers, *Electronics*, March, 1946, pp. 92, 106 (abstract only); T. M. Davis and Emerick Toth, Design of Communication Receivers for the Naval Service with Particular Consideration to the Very-high-frequency and Ultra-high-frequency Ranges, *Proc. IRE*, February, 1947, pp. 201-207; M. J. O. Strutt, "Modern Ultra- and Extreme-shortwave Reception," Van Nostrand, 1947; S. N. Van Voorhis, "Microwave Receivers," McGraw-Hill, 1948.

man-made interference, and (3) the necessity of handling greater band widths. The increase of the input conductance causes lower voltage gains in the coupling circuits. The reduced atmospherics and interference make weaker signals potentially usable. The greater band widths mean that sources of noise in the resistors and tubes of the receiver will produce greater noise power. The result of all these factors is that noise originating in the receiver is a major problem; in fact the design of front ends for receivers for these frequency ranges is largely a matter of obtaining adequate signal-to-noise ratio.¹

At increasingly higher frequencies, the input conductance of tubes, when operated in the grounded-cathode manner, becomes greater as a result of two causes. One is the presence of inductance in the cathode lead which is common to the input and output circuits of the tube. This is often minimized by the provision of two separate leads from the cathode to separate pins in the base of the tube, one of these being used for the input circuit and the other for the output circuit. Another cause of increasing input conductance is the transit-time effect resulting from the fact that the signal potential on the control grid changes appreciably in the time required for electrons to go from the cathode space charge to the grid, resulting in the signal on the grid doing work on the electrons and thus abstracting energy from the source of the signal and increasing the input conductance of the tube. The conductance due to both the lead-inductance and transit-time effects is proportional to g_m and to f^2 . Tripling the frequency causes the input conductance to increase to nine times its previous value. Thus a type 57 pentode was found experimentally to have an input conductance of 50 μ mhos at 30 Mc but 670 μ mhos at 100 Mc.

Since the noise power produced by tubes increases with the band width, the narrower the band, the higher the frequency limit at which any particular type of tube will work. A given tube will work satisfactorily as an oscillator at frequencies considerably above its limit as an r-f amplifier.

Superregenerative detectors for operation at 30 to 3,000 Mc proved serviceable in various wartime equipment.

Receivers for the frequency ranges now being considered have i.f.s from 5 to 200 Mc.² Values of 30 and 60 Mc are widely used in radars. The i-f band widths in various services have values up to 25 Mc. The wider band widths are obtained by staggered tuning or by the use of special feed-back arrangements. Higher i.f.s up to 300 Mc are being studied.

154. Use of Lumped-constant Circuits. It is often found that satisfactory tuning control can be obtained by switching in fixed wide-band antenna circuits and r-f tuned circuits upon change of band and varying only the oscillator frequency in tuning from station to station.³ In a receiver covering a band of relatively wide coverage and having low i.f., this technique would lead to objectionable image interference. However, if the limitations are borne in mind, this technique may be found a valuable method. It has been widely employed.

The circuit between an antenna and a conventional grounded-cathode r-f stage can advantageously be given a slight overcoupling in the interest of improved signal-to-noise ratio. This increases the skirt selectivity and thereby removes some noise in the outer frequency regions where no signal components exist.

Careful design of a 240-deg split-stator variable capacitor⁴ and its fixed inductor made possible a single-band coverage of the range from 200 to 400 Mc in a four-gang superhet having a two-circuit antenna coupling, a 956 acorn r-f stage, a tuned interstage coupling, a 955 or a 6F4 oscillator, and a triode mixer. Values of loaded Q from

¹ Herold, E. W., An Analysis of the Signal-to-noise Ratio of Ultra-high-frequency Receivers. *RCA Rev.*, January, 1942, pp. 302-331. Young, V. J., Understanding Microwaves. John F. Rider, Publisher, 1946. Note pp. 263-267.

² Lehenbaum, Matthew T., Wideband I-f Amplifiers above 100 Mc, *Electronics*, April, 1947, pp. 138-141.

³ As an example in amateur practice, see B. C. Barbee, A Compact and Inexpensive Superhet for 144 Mc. *QST*, October, 1947, pp. 33-36.

⁴ Davis and Toth, *loc. cit.*

45 to 65 per circuit were realized. The r-f stage gave an output-input voltage ratio which, with the 956 tube, averaged less than unity over the frequency range. However, the design satisfied the naval shipboard requirements for very low oscillator radiation.

When conventional coils and capacitors are applied at higher frequencies, a point is reached where the capacitance has been reduced to only that inherently present in the tube, socket, wiring, and coil, and the inductance is as small as possible for the required L/C ratio. Upon replacement of the tube, therefore, there will generally be a different input capacitance, the difference being a large enough fraction of the total circuit capacitance to render inaccurate a calibrated tuning dial or to disturb a unicontrol arrangement. Reduction of the tuning capacitance to where it consists only of the inherent minimums also has the objection that if a.v.c. is applied to the stage, the resulting variation of tube input capacitance may cause excessive detuning.

Lumped-constant tuning, with a discrete coil but only distributed capacitance, has been widely used in military equipment operating up to 200 Mc, and the technique has some possibility of usefulness above this. The limitations can be largely obviated at the expense of avoiding unicontrol tuning and avoiding a.v.c. The tuning is done inductively by inserting in the coil a sleeve of silver-plated metal to reduce the inductance to the desired degree, or by increasing the inductance by similarly employing a core of a suitable grade of compressed powdered iron. For covering the widest frequency range a composite core with the iron at one end and silver plating at the other can advantageously be used. Typical maximum coverage with these lumped-constant techniques is a 1.5:1 frequency range.

A unicontrol design with a three-gang variable inductor described by White¹ utilizes cylindrical cores of powdered iron over a frequency range from 110 to 135 Mc. The unicontrol is unusual in providing one dial with a step for each megacycle of frequency, the other dial serving for band spread or vernier control.

A sensitivity of more than 96 db (less than $15 \mu v$) for a 6-db value of the signal-to-noise ratio was achieved in one wartime receiver having separate-dial tuning control from about 160 to 200 Mc. This design used combination iron and silver slugs and the 956 as r-f amplifier. A.v.c. was not employed.

Another available design for frequencies up to 200 or 300 Mc is the grounded-cathode triode in a push-pull circuit. Laboratory work by William F. Bailey (Hazeltine) has shown that a 6J6 double triode can be cross-neutralized in the conventional transmitter fashion and good gain and signal-to-noise ratio thus obtained. However, the method has the drawback that the stage may require readjustment of the neutralization when a tube replacement is made.

155. Triodes in Grounded-grid Circuit. For signals from about 100 Mc, advantages may be found in the grounded-grid triode² as an r-f amplifier, especially if the receiver is intended for wide-band operation. In the grounded-grid amplifier the high side of the input connection is made to the cathode and the low side to the grid, which is grounded. The output is taken between plate and ground in the normal manner. Analysis shows that the input and the output currents are about equal, so that the amplifying action of the tube can be considered as reproducing the input current but at a higher voltage and out of a higher impedance.

The gain of a grounded-grid tube is

$$\text{Voltage gain} = \frac{R_L(\mu + 1)}{R_p + R_L}$$

where R_L = load resistance, ohms

μ = amplification factor of tube

R_p = internal plate resistance of tube, ohms

¹ White, S. Y., Very-high-precision Tunable Receiver for VHF, *Tele-Tech*, April, 1947, pp. 48-51, 106-107.

² Jones, M. C., Grounded-grid Radio-frequency Voltage Amplifiers, *Proc. IRE*, July, 1944, pp. 423-429. Dishal, M., Theoretical Gain and Signal-to-noise Ratio of the Grounded-grid Amplifier at Ultra-high Frequencies, *Proc. IRE*, May, 1944, pp. 276-284.

If the load resistance is much less than the internal tube resistance and if μ is much greater than unity, the gain simplifies to $R_L g_m$.

The value of the grounded-grid stage lies essentially in the fact that the input and the output circuits are fairly well shielded from each other, as in a screen-grid

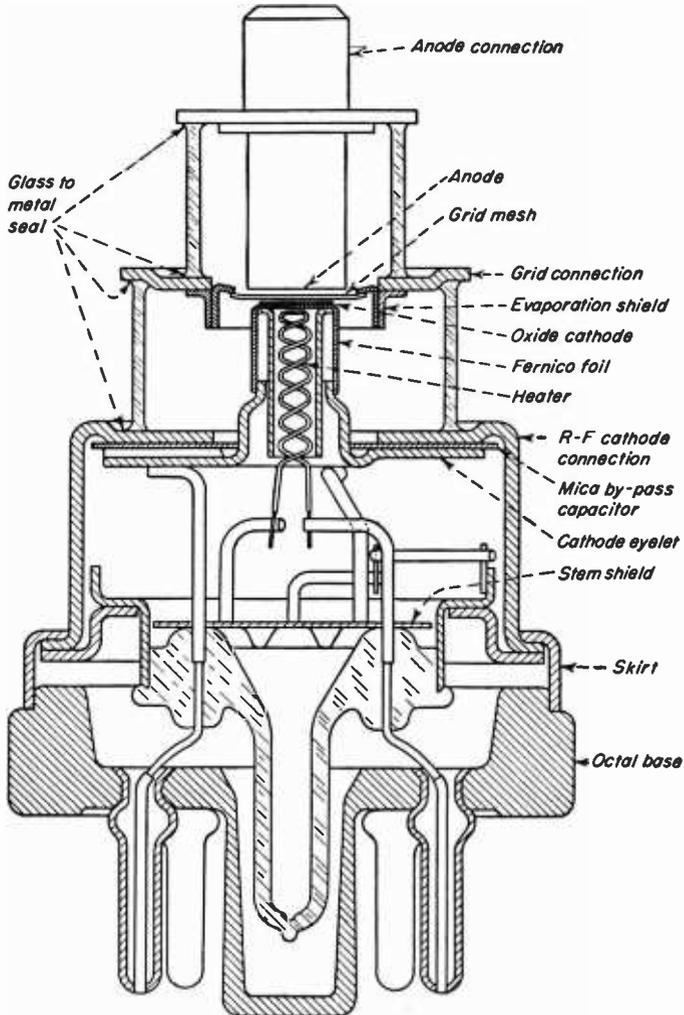


FIG. 36. Construction of type 2C40 lighthouse triode designed for grounded-grid operation. Over-all height is $2\frac{9}{16}$ in.

tube, while at the same time the low noise level characteristic of a triode is retained. Another advantage, which is important, is that the grounded-grid stage has an input conductance which is fairly constant with frequency, being

$$g_i = \frac{\mu + 1}{R_p + R_L}$$

If μ is much greater than unity and if the load resistance R_L matches the internal tube resistance R_p , the input conductance is $\mu/2R_p$, or $g_m/2$. If the load resistance is small compared with R_p and μ remains much greater than unity, the input conductance is μ/R_p , or g_m . Since the input conductance of a pentode increases steadily with frequency, there is a frequency above which the grounded-grid triode has lower input conductance than the pentode.

Slightly less than critical coupling into a grounded-grid stage gives the best signal-to-noise ratio.

The 6J4 miniature triode is useful for grounded-grid operation and is rated up to 500 Mc for this service. It has a low value of plate-to-cathode capacitance so as to reduce the feedback from output to input circuit inside the tube. A striking application of this tube was made in the radar receiver used to obtain echoes from the moon at 111.5 Mc. This receiver achieved a noise figure of only 5 db for the front end of the set.

For tuning in the general region of 300 to 1,300 Mc, especially over limited ranges, transmission lines¹ can be used. A $\lambda/4$ line shorted at the far end acts as a parallel tuned circuit at the near end. Other lengths of shorted line are available for use as high- Q inductors or capacitors. Carefully made transmission lines can be constructed to have the desired reactive characteristics and to have low losses. Balanced parallel-rod lines can be used up to 600 or 700 Mc, and coaxial types above this.

Transmission-line tuning over the range of 505 to 525 Mc was used in the Model YJ series of radar beacons during the war. A short length of parallel-rod type of line with an adjustable shorting bar was employed. This was the variable tuning element between the antenna and a 955 acorn triode acting as a superregenerative detector.

156. Use of Lighthouse Triodes in Grounded-grid Circuit. As signals above 400 or 500 Mc are to be received, it becomes advantageous to change from the conventional type of tube to the lighthouse version,² and usually at the same time to employ a coaxial type of transmission line rather than the parallel-rod type for tuning.³ The lighthouse tube (which is sometimes called the "planar" type from the fact that the active surfaces of the elements are planes) is constructed with flanges or belts around the edge so that large low-inductance connections can be made to the coaxial type of transmission line. The 2C40 tube, shown in Fig. 36, is rated up to 1,200 Mc as an r-f amplifier, and experience has shown that this range is obtainable for narrow-band operation. For wide-band service, the limit of usefulness is 800 or 900 Mc. The tube is rated to 3,370 Mc as a c-w oscillator. With an amplification factor of 36 and a plate resistance of 6,000 ohms, the transconductance is 6,000 μ mhos. The plate-cathode capacitance is 0.04 μ mf.

The short lengths of these coaxial transmission lines and the large circular contacts required for connection to the lighthouse tubes give equipment of this type an appearance which has led to the term "plumbing" as the usual informal name. The characteristics and name continue to apply with preselectors and local oscillators of the transmission-line or wave-guide type up into the s-h-f range.

157. Butterfly Tuners.⁴ Tuners of the butterfly type, such as shown in Fig. 37, are continuously adjustable over a wide frequency range, designs of moderate size achieving ratios as great as 5:1 in frequency and larger designs achieving greater ratios. These tuners have been made and used at frequencies as low as 40 Mc and as high as 1,100 Mc or more. The construction is somewhat similar in appearance to a variable capacitor, with the exceptions that (1) the circuit inductance is built in, consisting chiefly of the circular-strap portions of the stator plates, (2) both connections are made

¹ Meagher, J. R., and H. J. Markley, "Practical Analysis of Ultra-high-frequency," RCA Service Company, 1943. This pamphlet gives a compact clear treatment of transmission lines and wave guides.

² McArthur, E. D., Disk-seal Tubes. *Electronics*, February, 1945, pp. 98-102.

³ Petrich, R. O., A Wide-band 550-megacycle Amplifier. *Proc. IRE*, November, 1947, pp. 1371-1374.

⁴ Karplus, Eduard, Wide-range Tuned Circuits and Oscillators for High Frequencies. *Proc. IRE*, July, 1945, pp. 426-441. A shorter account by the same author entitled The Butterfly Circuit appears in *Gen. Radio Experimenter*, October, 1944.

to the stator, whence (3) there are no sliding contacts. When the rotor is turned toward increased meshing, the frequency is lowered by (1) increased inductance, resulting from removal of the conducting rotor plates near the inductive straps, and (2) increased capacitance in each of the two series capacitors (one-terminal-to-rotor and rotor-to-other-terminal) constituting the capacitive element of the tuner. The rotor turns through an angle of 90 deg for its full frequency range.

One design of butterfly tuner covering the range from 220 to 1,100 Mc has the following characteristics:

Frequency, Mc.....	220	1,100
Inductance, μ h.....	0.011	0.004
Capacitance, μ mf.....	48	5
Q	650	300
Impedance, ohms.....	9,800	8,600

This design has a diameter of $2\frac{1}{2}$ in. It has 0.04-cm clearance between stator and rotor plates as they mesh, so that high voltage and power cannot be handled. It has five rotor and six stator plates.

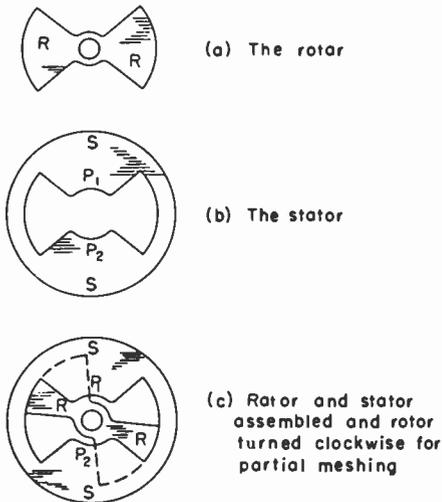


FIG. 37. The butterfly tuner. Notation: P_1 and P_2 = points for electrical connections; R = rotor; S = stator.

results in lower gain or lower signal-to-noise ratio than without the stage.² Klystrons will produce amplification at these frequencies and are useful as transmitting amplifiers, but they are too noisy for use in receivers.

The receiver for frequencies above this point, therefore, has as its first elements the antenna, a passive coupling circuit, and a silicon-crystal mixer for converting the signal to i.f.

As an example of a design of this kind, operating in the band from 950 to 1,150 Mc, the Model OD (Hazeltine) receiver³ uses a type 1N25 crystal mixer, followed by a

¹ Karplus, *loc. cit.*

² In connection with the maximum frequency at which a lighthouse triode can be used to advantage, reference may be made to N. T. Lavoo, Transmittance and Input Conductance of a Lighthouse Triode at 3,000 Megacycles, *Proc. IRE*, November, 1947, pp. 1248-1251.

³ This receiver is described in Report 28 of the Combined Research Group of the Naval Research Laboratories, which is available from the U. S. Dept. of Commerce as Publication PB-L-79529, either a photostat or a microfilm copy.

Butterfly tuners are suitable for use where wide frequency ranges must be handled. In design for such service, the size and number of plates are made as large as possible while still getting the necessary top frequency and freedom from spurious modes of operation. The necessary i.f. is then obtained by reducing the air-gap clearance as far as required.

A tuned circuit of cylindrical shape⁴ for u-h-f use has an approximately constant Q over its frequency range. For example, a cylinder of 1 in. diam and 2 in. length tunes from 500 to 1,000 Mc with a Q of 1,200. This type of circuit will not cover so wide frequency ranges as the butterfly designs. Care is advisable regarding performance over whatever temperature range is required.

158. Technique for Frequencies Where R-f Stage Cannot Be Used.

At some point in the frequency band between 500 and 3,000 Mc, it is found that the use of an r-f stage in a receiver

60-Mc 6-stage i-f amplifier using 6AK5 pentodes. This design has an over-all pass band of approximately 8 Mc and achieves a sensitivity of about 102 db ($8 \mu\text{v}$) at the point where a 6-db signal-to-noise ratio prevails. An image rejection of 25 db is obtained. Resistance feedback loading is used on five of the i-f stages in order to obtain the required band width. A coaxial-cavity plumbing is used for signal tuning of the circuit between the antenna and the crystal and also for housing and tuning the 2C40 lighthouse local oscillator. Preset tuning to any one of a number of fixed frequencies is included, the method being very similar to that in the receiver of the distance-measuring equipment described in Sec. 70.

The reflex klystron¹ is generally used as the oscillator for receivers in the s-h-f range. The likelihood of requiring a.f.c. is apparent when it is noted that a band width of 10 Mc at a carrier frequency of 10,000 Mc amounts to only 1 part in 1,000. A.f.c. is therefore generally provided, the control system operating by placing on the reflex electrode of the klystron a suitable voltage so as to keep the i.f. of the receiver at its proper value.

For wide-band reception, such as for narrow sharp pulses, in this frequency range, extensive design procedures from the television and radar² arts are available and are described in Secs. 194 to 196.

Plumbing between the antenna and the crystal converter has been designed with both loop and aperture types of coupling. Some recent experimental work³ has indicated that the aperture type is preferable because it can be designed to have almost zero self-reactance and to have the apertures flush with the cavity walls, thus producing a minimum effect on the field and on the tuning of the cavities. As to the tuning cavities themselves, it was reported that the coaxial type operating on the principal (transverse-electromagnetic) mode has proved quite satisfactory. With tuners having three cavities, it was found that an unloaded Q of 1,000 is required for each cavity to get a 1 per cent band width and keep the total dissipation loss of the tuner under 1.5 db. An experimental tuner of this kind was adjustable over the frequency range from 7,000 to 11,000 Mc using cavities of $3\lambda/4$ and tuning by motion of the center conductors.

Considerable interest has been shown in the traveling-wave tube described by Kompfner in England⁴ and Pierce and Field⁵ in the United States. An experimental tube of this type made at the Bell Laboratories furnishes substantial amplification over the very wide band width from 3,200 to 4,000 Mc.

SUPERREGENERATIVE RECEIVERS

Superregeneration offers high gain, such as 100 db, in a single tube. Fundamentally a method of amplification, it is easily adapted to performing a-m detection also and has been very frequently used with this additional feature. On account of the high gain obtainable, it is possible to make a set of high sensitivity with only a few tubes. Superregeneration, described by Edwin H. Armstrong in 1922,⁶ was little used except in amateur v-h-f receivers until the Second World War when it was employed extensively in military equipment. Subsequently it was given further careful study.⁷

¹ Pierce, J. R., and W. G. Shepherd, Reflex Oscillators, *Bell System Tech. J.*, July, 1947, pp. 460-681. Harrison, A. E., "Klystron Tubes," McGraw-Hill, 1947. McNally, J. O., and W. G. Shepherd, Reflex Oscillators for Radar Systems, *Proc. IRE*, December, 1947, pp. 1424-1435.

² Schneider, Edwin G., Radar, *Proc. IRE*, August, 1946, pp. 528-578; see section on receivers, pp. 551-556. Miller, S. E., Considerations in the Design of Centimeter-wave Radar Receivers, *Proc. IRE*, April, 1947, pp. 340-351.

³ Petrich, Raymond O., Coupled Circuits Used as Tunable Band-pass Filters in the Ultra-high-frequency and Microwave Regions, paper presented at IRE Convention, New York, March, 1947.

⁴ Kompfner, R., The Traveling-wave Tube as Amplifier at Microwaves, *Proc. IRE*, February, 1947, pp. 124-127.

⁵ Pierce, J. R., and L. M. Field, Traveling-wave Tubes, *Proc. IRE*, February, 1947, pp. 108-111. Pierce, J. R., Theory of the Beam-type Traveling-wave Tube, *Proc. IRE*, February, 1947, pp. 111-123.

⁶ Armstrong, Edwin H., Some Recent Developments in Regenerative Circuits, *Proc. IRE*, August, 1922, pp. 244-260.

⁷ Hazeltine, Alan, D. Richman, and B. D. Loughlin, Superregenerator Design, *Electronics*, September.

159. General Principles. The superregenerator tube is provided with ample feedback so as to be capable of oscillation at the desired r.f. It is also provided with a periodic quench voltage by which an oscillation which has built up is stopped for a time, after which the quench is removed and an oscillation builds up again. In this way the circuit is prevented from going into permanent oscillation; instead it engages in a periodic building up of transient oscillations whose average value depends upon the signal voltage existing in the circuit when each build-up starts. Thus amplification is produced. Superregeneration can therefore be considered as a sampling process by which samples of the signal are taken at quench-frequency intervals and much amplified. This process is sufficient to recreate the envelope, if the highest frequency component of the signal is less than half the quench frequency.

In case no signal is present, fluctuation noise will start the oscillations, and this noise will therefore be amplified by the superregenerative action. However, impulse noise, which occurs at discrete instants, generally does not coincide with the sensitive periods of the superregeneration, so that most of the impulses are suppressed.

For purposes of analysis and design, the action of the superregenerative tube can be considered as presenting to the tuned circuit a negative resistance, or the equivalent negative conductance, during the build-up time. Conversely, during the quench interval the circuit has loss corresponding to a positive conductance. The result is that the conductance of the tuned circuit varies with time at the quench frequency. H-f oscillations build up during the period of negative conductance and decay during the following period of positive conductance. The conductance-time characteristic is one of the fundamental quantities determining the performance. Another important quantity is the total tuned-circuit capacitance.

The quench voltage can be provided by a separate oscillator, in which case the circuit is described as having "separate quench." Certain operating features are obtainable only in this way. Often, however, the superregenerative tube is provided with a suitable grid leak and capacitor so that its oscillations block periodically, thus enabling the tube to serve as its own source of quench voltage. Such operation is described as "self-quenched."

There are two distinct modes¹ of superregenerative operation: the linear and logarithmic.

In the linear mode the quench frequency and other factors are so proportioned that the quench voltage acts to stop the building up of oscillations while they are still rising; *i.e.*, the oscillations do not reach the maximum amplitude at which the tube would operate if it were in a steady state of oscillation. The linear mode is given this name because the amplitude reached is proportional to the signal amplitude at the start of the build-up period; in this way there is a linear mathematical relation between output and input. Separate quench is required for the linear mode.

In the logarithmic mode the oscillations are allowed to build up to the maximum amplitude of which the tube is capable before the quench voltage is applied. Build-ups starting from various signal strengths are, however, not identical, because one started by a strong signal has a larger value throughout the entire build-up period and reaches the maximum value sooner than one starting from a weak signal. The modulation-frequency signal obtained upon detection is found to be proportional to the logarithm of the input signal amplitude, so that this type of operation is called the "logarithmic mode." This mode can be obtained with either separate or self-quench.

A logarithmic-mode superregenerator can easily be adapted to perform a-m detection in addition to superregenerative amplification, by providing an impedance in either the grid or plate circuit across which the m-f voltage can be developed.

In self-quenched operation the quenching action can occur only as a result of the

1948, pp. 99-102. Bradley, William E., Superregenerative Detection, *Electronics*, September, 1948, pp. 96-98.

¹ Frink, F. W., The Basic Principles of Superregenerative Reception, *Proc. IRE*, January, 1938, pp. 76-106.

oscillations building up to the maximum amplitude which is possible under the particular conditions. Such operation is therefore always of the logarithmic mode. With a strong signal the maximum amplitude is reached sooner, the quench voltage comes on sooner, and after the usual quiescent period, the next build-up of oscillations is sooner—all in comparison with the action for a weak signal. Therefore, the quench frequency varies in self-quenched operation, the value rising and falling with the amplitude of the signal.

If a superregenerative receiver oscillates at or near the signal frequency, it may be an objectionable source of radiation to other nearby receivers tuned to receive the same transmission. This radiation can be reduced by using an untuned shielded r-f stage ahead of the superregenerator.

The wave form of the quench voltage may be of several different shapes: sine-wave, rectangular, trapezoidal, saw-tooth, etc. The resulting curve of circuit conductance as a function of time will vary accordingly, and the performance may thus be considerably affected. As an example, selectivity is better for sine-wave quench than for rectangular-wave quench.

The table on page 866 gives the chief characteristics of typical superregenerative circuits used in recent years.

160. The Linear Mode. Operation in the linear mode requires a separate quench. The gain in this case and the selectivity for either the linear or the logarithmic mode¹ are obtainable from a knowledge of the circuit-conductance curve and the total circuit capacitance. The superregenerative gain is $4.35a/C$ db, where a is the area of the negative-conductance portion of the conductance-time curve (Fig. 38) and C is the total circuit capacitance.

Example: If a is 325×10^{-12} mho-sec and C is $20 \mu\text{mf}$, the superregenerative gain is 70 db. This is to be added to the nonregenerative gain in the input coupling system and to the regenerative gain. Thus, if the nonregenerative gain is 20 db and the regenerative gain is 14 db, the total gain would come to an over-all figure of 104 db.

The band width for either linear or logarithmic operation, measured at 8.7 db (1 neper) down, is given by

$$\frac{1}{\pi} \sqrt{\frac{g'}{C}}$$

where g' is the rate (expressed as a positive quantity) at which the conductance is changing when the sign of the conductance changes from plus to minus, and C is the total capacitance of the circuit. This formula can be used with saw-tooth, trapezoidal, and sine-wave quenches, but not with square-wave quench; *i.e.*, the value of g' must not be too large. In particular, this formula is useful when the value of g' is less than $G^2/4C$, where G is the absolute value of the conductance at the end of the straight-line portion of the conductance-time curve, as shown in Fig. 38.

Example: If g' is 200 mhos per sec and C is $20 \mu\text{mf}$, the 8.7-db band width is 1 Mc. The value of $G^2/4C$ is 405, so that the criterion for the applicability of the formula is satisfied.

¹ Macfarlane, G. G., and J. R. Whitehead, The Theory of the Superregenerative Receiver Operated in the Linear Mode, *J. IEE (London)*, Part III, May, 1948, pp. 143-157. This paper gives some of the formulas which were developed by Harold A. Wheeler and incorporated in internal Hazeltine reports 1457-W of Jan. 5, 1943 and 1457-AW of Feb. 16, 1943. These reports are believed to have given the first simple analysis of the selectivity of the superregenerative circuit.

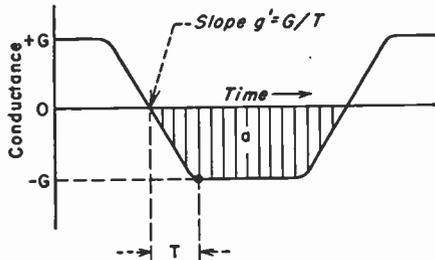


FIG. 38. Trapezoidal quench wave showing quantities which determine the sensitivity and selectivity of a superregenerator.

Characteristics of Various Superregenerative Receivers

Service	Carrier range, Mc	Type of signal	R-f stage used	Mode	Quench frequency	Separate detector used	Remarks
Amateur practice ^a	60-220	A-m voice and telegraph	In some cases	Log, self-quenched	20-250 kc	No	Sometimes preceded by frequency converter so that superregeneration is at i.f. of about 25 Mc
Reception of radiosonde signals	25-280	A-m voice and m-c-w telegraph; also a-m signals for meteorological quantities	Yes	Log, self-quenched	50 kc	No	National Co. Model 1-10A
Control of model planes ^b	56-60	A-m pulses like telegraph dashes	No	Log, self-quenched	Not stated	No	Uses Raytheon RK-42 hard triode or RK-62 gas triode
Military beacons and identification equipment ^c	150-240	Pulses of 2-10 μ sec	In some cases	Linear	300-600 kc	Yes	
	510-520	Pulses of 2 μ sec	No	Linear	600 kc	Yes	
Remote control of missiles ^d	50-100	A-m tones	Yes	Log, separate quench	50-100 kc	No	Selectivity improved in redesign by use of lower quench frequency and amplitude
SCR-194 and SCR-195 military walkie-talkies	28-65	A-m voice	No	Log, self-quenched	15 kc	No	Crystal included as frequency standard
Model RC-57 for remote control of target planes ^e	68-73	A-m tones	Yes	Log, self-quenched	Not stated	No	
Hazeltine FreModyne f-m broadcast receiver	88-108	F-m sound	Usually not	Log, self-quenched	30 kc	No	Tube converts to i.f. of about 22 Mc and superregenerates at this frequency
Military transceiver ^f	2,700	A-m voice	No	Mode not stated, separate quench	100 kc	No	Uses 2C40 lighthouse tube as superregenerative detector when receiving
Special purpose	1,000-3,000	Pulse	No	Linear	1,000-2,000 kc	Yes	
Citizens-band transceiver ^g	460-470	A-m voice	No	Log, self-quenched	Not stated	No	Square-wave quench
				Mode not stated, separate quench	Not stated	No	

^a "Radio Amateur's Handbook," pp. 316-329, American Radio Relay League, West Hartford, Conn., 1946.

^b *Electronics*, February, 1948, pp. 81-83; *QST*, March, 1940, pp. 24-27, 86, 88

^c Wood, K. A., 200-Mc/s Radar Interrogator-beacon Systems, *J. IEE (London)*, Part III A, No. 2, pp. 481-495, 1946.

^d *Electronics*, March, 1946, pp. 154-157.

^e *Electronics*, December, 1946, pp. 89-90.

^f *Electronics*, September, 1946, pp. 104-105.

^g *Electronics*, January, 1948, pp. 92-96.

Further information regarding band width is obtainable from the fact that the selectivity curve has the shape of a probability curve between the nose and a value of attenuation equal to $2.2 GT/C$ db, where G and T ($=$ time) have the values at the point indicated by a dot on the conductance curve of Fig. 38. Assuming, in the typical example, that the value of T is $0.9 \mu\text{sec}$, the formula gives 18 db. In other cases this may amount to 40 to 80 db.

A probability curve has a shape corresponding to the selectivity of a large number of isolated single-tuned circuits in cascade. A superregenerative receiver can, therefore, have a selectivity curve with very steep sides. By comparison with the selectivity obtainable in conventional multicircuit receivers, the probability curve is somewhat broad near the resonant frequency; yet it has good skirt selectivity, which means good rejection of unwanted signals outside the pass band of the receiver. For the best selectivity a saw-tooth quench wave can be used with the gradual change from positive to negative conductance and the rapid change in the opposite direction. This decreases the value of g' and increases the value of T .

Linear superregeneration was widely used in the transponders of identification and radar-beacon equipment during the Second World War, and listings for these services have been given in the accompanying table. The YJ series of radar beacons operated at 176 and at 515 Mc, switching between these two frequencies about 18 times per sec. A common 600-kc oscillator for separate quenching was used. The superregenerative tube for the 176-Mc channel was a type 7193 (similar to the 2C22), and that for the 515-Mc channel was an acorn 955. Separate miniature 9004 diodes were used as detectors on the two frequency ranges.

Automatic gain stabilization (AGS) was included in these beacons to maintain reception despite changes of antenna impedance, line voltage, tube aging, and the effects of high humidity. The system operates by detecting the quench-frequency pulsation of the r-f signal in the superregenerative circuit, amplifying it, rectifying it, and using the resulting d-c voltage to control the superregenerative tube. This maintains the quiescent noise output almost constant, in addition to keeping the general operation within good limits.

The sensitivity of a linear-mode superregenerator increases as the amplitude of the quench voltage increases, but the quench amplitude must not be great enough to overload the tube. Sensitivity also increases as the frequency of the quench voltage is lowered. Selectivity increases, *i.e.*, band width decreases, for both lower frequency and lower amplitude of quench voltage. It is seen therefore that sensitivity and selectivity improve with lower values of the quench frequency. The requirements of sensitivity and selectivity with regard to amplitude are however opposed, low amplitude favoring selectivity and high amplitude favoring sensitivity.

161. Logarithmic Mode with Separate Quench. Theoretical and experimental studies of superregeneration of the logarithmic mode with separate quench have been made.¹ There is considerable amplitude distortion with the logarithmic mode because a large increase of signal causes only a slight increase in the output due to the logarithmic output-input relationship. This is advantageous in reducing the effect of impulse noise, but produces considerable distortion in a-m reception. For headphone listening to speech signals the modulation under these conditions should therefore not exceed about 80 per cent.

162. Self-quenched Circuits in A-m Reception. All self-quenched circuits operate in the logarithmic mode, and the quench frequency is a function of the signal amplitude. Also, all quench groups of oscillations are substantially identical in shape.

The self-quenched superregenerator has been extensively used in amateur practice,² triodes such as the 6J5, 7A4, 9002, and 6C4 being employed with grid leaks of 2 to 10 megohms and grid capacitors of 50 to $100 \mu\text{f}$. The 6F4 is also very suitable. Quench

¹ Ataka, H., On Superregeneration of an Ultra-short-wave Receiver, *Proc. IRE*, August, 1935, pp. 841-884. Strafford, F. R. W., The Superregenerative Detector: An Analytical and Experimental Investigation, *J. IEE (London)*, Part III, No. 21, January, 1946, pp. 23-28.

² "Radio Amateur's Handbook, 1946," pp. 316-329, American Radio Relay League, West Hartford, Conn.

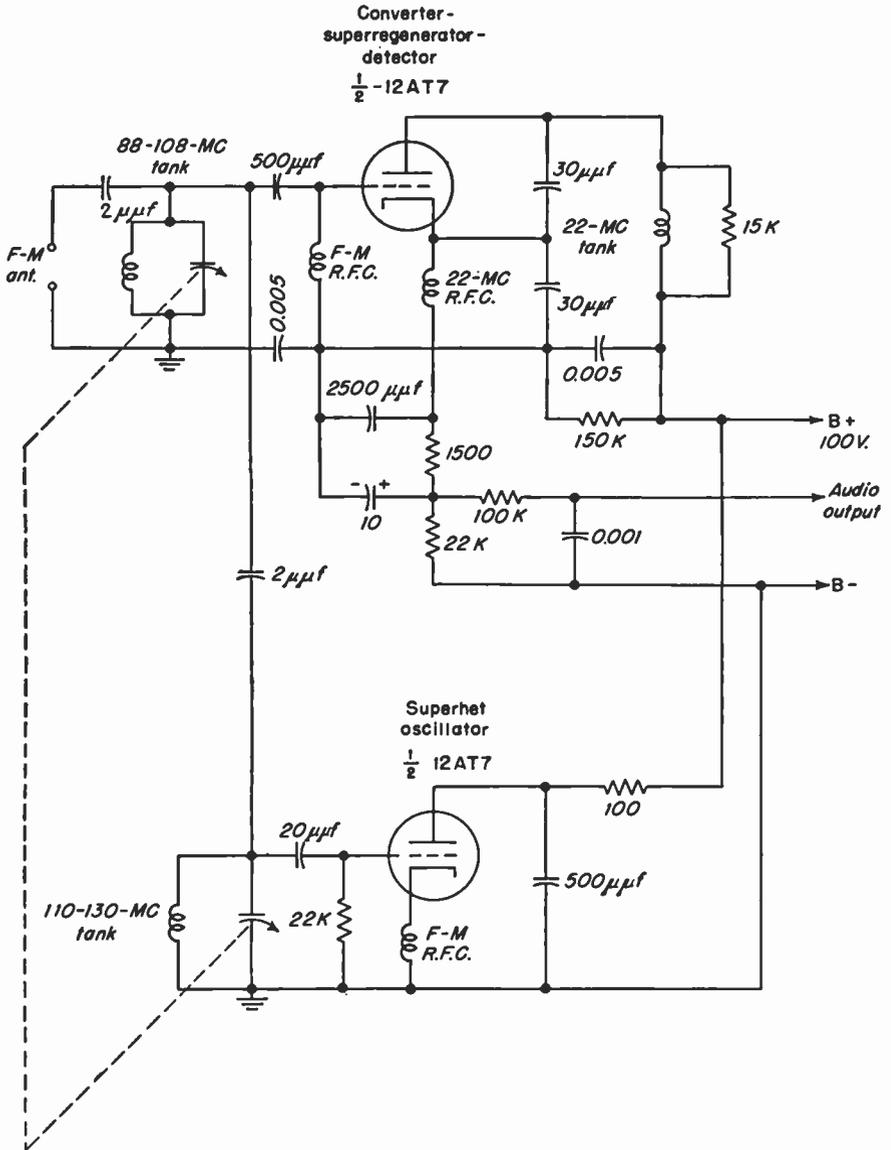


FIG. 39. The Fremodyne superregenerative-superheterodyne circuit.

frequencies have varied from 20 to 250 kc as the center value, with frequencies of 150 to 250 kc recommended for signals at 60 Mc and higher in the v-h-f range.

One refined amateur superregenerator¹ includes an untuned 6AK5 r-f stage and achieves a substantial reduction of radiation.

¹ Tilton, E. P., A Non-radiating Superregenerative Receiver for Two Meters, *QST*, February, 1946, pp. 53-56, 108.

163. Superregenerative Circuits in F-m Reception. By detuning a superregenerative receiver to one side of the selectivity curve, f-m signals can be converted into a-m form in addition to being amplified. The same tube can furnish the self-quench, if desired, and also detect the a-m signal produced by the side tuning so as to deliver audio output.

A versatile circuit of Bernard D. Loughlin¹ combines superregenerative, superheterodyne, and detecting features in one double-triode tube. The circuit converts the v-h-f signals received on the f-m antenna into an audio signal which is large enough to operate the conventional audio amplifier of a receiver. One triode serves as the local oscillator necessary for frequency conversion. The other triode performs four functions, operating as (1) a superheterodyne mixer producing an i.f. of about 22 Mc, (2) a high-gain superregenerative i-f amplifier, (3) a converter from f.m. to a.m., and (4) an a-m detector delivering audio output. The conversion of the f-m signal to a.m. in step (3) is accomplished by side tuning on the i-f selectivity curve. The superregeneration is of the self-quenched logarithmic mode with a quench frequency in the neighborhood of 30 kc. The circuits associated with this dual triode are given in Fig. 39.

The use of the superheterodyne principle in this circuit greatly reduces signal-frequency radiation compared to a conventional superregenerator, making it somewhat less than produced by many conventionally designed medium-priced f-m receivers. A special automatic stabilizing arrangement makes unnecessary any manual regeneration control and also permits a quench-wave shape to be obtained which gives good selectivity, good audio output, and quite linear f-m detection.

The usable f-m sensitivity of this circuit is represented by the quieting sensitivity, defined as the signal level required to give an audio output for 30 per cent modulation which is 30 db above the background noise. For a typical set this sensitivity is 74 db (200 μ v) with a 300-ohm dummy antenna. A signal weaker than 74 db can be heard but at a correspondingly poorer signal-to-noise ratio. An 83-db (70- μ v) signal gives approximately 20-db signal-to-noise ratio.

The selectivity is shown in Fig. 40 and is better than for many conventionally designed f-m receivers. It is sufficient for good rejection of local stations, particularly when the receiver is side-tuned on the appropriate side of the desired signal, *i.e.*, away from the interfering signal. The sides of the curve are practically straight lines; since the ordinates are decibels plotted linearly, the f-m detection by side tuning offsets the amplitude distortion associated with the logarithmic mode.

The superregenerative principle is applicable to the amplification of f-m signals, followed by separate detection. Studies of the frequency spectra in this case² show that the superregenerator produces a carrier with side bands, and in addition other groups of frequencies, separated at intervals of the quench frequency, each of these groups having a central "carrier" and signal side bands. This imposes a lower limit on the quench frequency; too low a value would cause more than one set of frequencies to be applied to the following f-m detector.

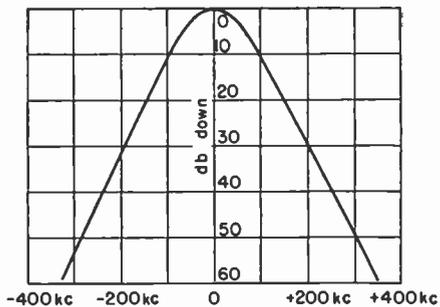


FIG. 40. Typical selectivity curve of FreModyne circuit.

¹ Hazeltine FreModyne FM Circuit, *Tele.-Tech.*, December, 1947, pp. 41, 85-86. FreModyne FM Receivers, *Electronics*, January, 1948, pp. 83-87.

² Kalmus, H. P., Some Notes on Superregeneration with Particular Emphasis on Its Possibilities for Frequency Modulation, *Proc. IRE*, October, 1944, pp. 591-600.

RECEIVERS FOR PULSE-TYPE COMMUNICATIONS

During the Second World War and the postwar period, there was an increasing realization that there are advantages in the use of pulses for general communication purposes,¹ such as multiplex telephony. Such a system utilizes the principle of sampling; *i.e.*, the signal wave to be transmitted is sampled at a frequency at least twice the highest component, so that the reproduced wave will be a satisfactory copy of the original.

164. Pulse Systems. Two of the several possible types of pulse systems are of importance: (1) pulse-time modulation (called PTM, or PPM for the alternative name pulse-position modulation) and (2) pulse-code modulation (PCM, also called pulse-count modulation). Both types use pulses of substantially rectangular shape and of constant amplitude. Since the sampling is at a frequency two or more times the highest signal component and since a number of harmonics are necessary to create the rectangular shape, the band width required by pulse systems is many times that needed in the conventional types of modulation. However, there are compensating advantages: (1) These systems are highly resistant to the effects of noise so that lower transmitting power can be used or a greater distance be obtained, and (2) in multiple operation where several messages are being sent simultaneously, nonlinear amplifiers can be used without danger of cross talk between the individual channels. It will be seen that these systems with pulses of constant amplitude have characteristics similar to telegraph practice where "regenerative" repeaters have long been used, the term "regenerative" referring to the restoring of the pulse shape in addition to the amplification in the repeater. On account of the simplicity of the wave shape, repeaters in pulse systems can produce such regeneration, so that the noise is substantially removed at each repeating point. In PTM and PCM systems the signal-restoring benefits realized in telegraph practice can be obtained also in the much faster operations required for the transmission of signals for voice, facsimile, etc.

165. Pulse-time Modulation. One method by which audio signals can be transmitted with pulses is by varying the position or time of successive pulses, each with respect to a reference pulse which occurs at a regular rate. Thus modulation is attained by displacing the signal pulse from its normal position or time by an amount proportional to the amplitude of the modulating signal. For example, if the pulse occurs later than normal, the audio amplitude may be positive; if the pulse occurs earlier, the amplitude will then be negative. It is possible to use a single reference pulse for numerous signal pulses, each signal pulse representing a different voice or other transmission so that multiplex operation is obtained.²

The effect of noise in a pulse-time system is limited to the extent to which it influences the time of occurrence of the signal pulse. As a result, the steeper the edges of the signal pulse, the less the influence of noise. Therefore, with increase of band width, which permits steeper slopes, noise has less effect on transmission. This system, therefore, obtains improved signal-to-noise ratio at the expense of greater band width.

The pulse-time method of modulation was used in the Signal Corps AN/TRC-5 and AN/TRC-6 radiotelephone equipments during the Second World War.³ Both were

¹ Pulse Communication, *J. IEE (London)*, Part IIIA, No. 11, pp. 83-106, March and April, 1947. This gives one long paper and ten summaries of papers presented in March, 1947. Bandwidth vs. Noise in Communications Systems, *Electronics*, January, 1948, pp. 72-75. How Rigid Is the Hartley Law? *Tele-Tech*, January, 1948, pp. 52-53.

² Deloraine, E. M., and E. Labin, Pulse Time Modulation, *Electronics*, January, 1945 pp. 100-104; also *Elec. Commun.*, 23, 91-98, 1944.

³ Kelleher, John J., Pulse-modulated Radio Relay Equipment, *Electronics*, May, 1946, pp. 124-129; describes both AN/TRC-5 and AN/TRC-6. Greer, William R., Pulse Modulating System, *Electronics*, September, 1946; pp. 126-131; describes AN/TRC-6. Black, H. S., J. W. Beyer, T. J. Grieser, and F. A. Polkinghorn, A Multichannel Microwave Radio Relay System, *Elec. Eng. Trans. Section*, December, 1946, pp. 798-806; describes AN/TRC-6. Lacy, Raymond A., Two Multichannel Microwave Relay Equipments for the United States Army Communication Network, *Proc. IRE*, January, 1947, pp. 65-70; describes AN/TRC-5 and AN/TRC-6 with accounts of use in the Second World War. Trevor, Bertram, Orville E. Dow, and William D. Houghton, Pulse Time Division Radio Relay, *RCA Rev.*, December, 1946, pp. 561-575; describes AN/TRC-5.

eight-channel multiplex systems, each channel having an audio range of 300 to 3,000 cycles. The AN/TRC-5 operated on one carrier between 1,350 and 1,500 Mc in one direction, and another carrier in this region in the opposite direction. The marker or reference pulse had a duration of 2 μ sec and occurred at a 10-kc frequency, *i.e.*, every 100 μ sec. The signal pulse had a duration of 0.4 μ sec, and each varied in position over a 7- μ sec range, *i.e.*, ± 3.5 μ sec from its normal position. To provide "guard bands" of time, each time channel in which the signal pulse was located had a duration of 11.1 μ sec.

The receiver of the AN/TRC-5 included a 1N21 crystal mixer, a 2C40 triode as local oscillator and had a 16-Mc i.f. The i-f band width at 3 db was 3 Mc. This receiver secured an image rejection of 28 db and a noise figure of 12 db.

The AN/TRC-6 operated in the s-h-f region of 4,350 to 4,800 Mc and used a 4- μ sec marker pulse at an 8-kc frequency. The signal pulse had a duration of 1 μ sec and a total range of 12 μ sec in a 15- μ sec time channel.

The AN/TRC-6 receiver included a crystal converter, a 2K22 klystron as local oscillator, and a.f.c. including a search feature. The i.f. was 60 Mc, and the band width was approximately 10 Mc. This receiver had very little image rejection and a noise figure of 20 db. The receiver included a "slicer," or double-clipper, type of circuit, which utilizes a horizontal slice through the pulse so that noise irregularities of small amplitude at the bottom of the pulse and also noise irregularities at the top of the pulse are eliminated. The height of this slice is only 5 per cent of the total amplitude of the pulse.

A gating system separated the various incoming signal pulses of the multiplex group into their respective channels; *i.e.*, all pulses are applied to all channels, but each channel is disabled at all times except when a-gating pulse enables it to receive the proper signal pulse.

After separation of the signal pulses into their eight individual channels, each is converted from PTM to pulse-width modulation by the use of a relaxation oscillator circuit so arranged that the signal pulse starts it into operation and the termination of the gate stops its operation. In this way, the length of time that the relaxation circuit operates depends on the position of the signal pulse, because this determines the beginning while the end is constant. The pulse-length-modulated signal thus obtained, when put through a 3,000-cycle low-pass filter, delivers the desired audio output without any higher frequency components resulting from the pulses.

A 24-channel PTM system developed by the Federal Telecommunications Laboratories,¹ required the transmission of modulating signal components up to 2.8 Mc. This system, suitable for use over a coaxial cable or over a radio link, utilizes pulses of 0.5 μ sec duration, two of these with a 1.3- μ sec separation constituting the marker pulse. Each signal pulse varies in position over a 2- μ sec range in a 5- μ sec time channel. Each of the signal channels is sampled 8,000 times per sec, corresponding to a 125- μ sec interval. This 125 μ sec is divided into 5- μ sec time channels to accommodate the 24 multiplex channels and the marker channel. The frequency of pulses, including those for all channels, is therefore $8,000 \times 25 = 200,000$ cycles, or 200 kc, so that the band width of 2.8 Mc includes the first 14 harmonics and permits a fairly good reproduction of the rectangular pulse shape.

The Federal PTM system includes two Cyclophon tubes, one of which is used as the commutator in the modulator of the transmitter. The other is used as separator and demodulator in the receiver. This tube is a modified cathode-ray tube in which the beam is whirled around in a circular path where it goes through successive apertures in an aperture plate and falls on individual "dynode" plates connected to the various channels. The system has been operated by radio² in the u-h-f range of 1,225 to 1,325 Mc. The receiver utilizes a 1N23B crystal mixer and a 2C43 oscillator. An image rejection of 72 db and a noise factor of 14 db are obtained. The i.f. is 30 Mc.

¹ Grieg, D. D., and A. M. Levine, Pulse-time-modulated Multiplex Radio Relay System—Terminal Equipment, *Elec. Commun.*, June, 1946, pp. 159-178.

² Grieg, D. D., and H. Gallay, Pulse-time-modulated Multiplex Radio Relay System—Radio-frequency Equipment, *Elec. Commun.*, June, 1947, pp. 141-158.

The band width at 1 db down is 8 Mc so that the 2.8-Mc maximum signal frequency, corresponding to 5.6 Mc for both side bands, is easily accommodated.

166. Pulse-code Modulation.¹ The method of sampling which characterizes all pulse-type communication systems is used in pulse-code modulation with the additional feature that, after a sample is taken, it is quantized; *i.e.*, a determination is made as to which one of a definite number of amplitude classes it falls into. The number of these classes is some power of 2, such as $2^4 = 16$, $2^5 = 32$, etc., the particular number depending on the design of the system.

However, for the purpose of explanation, assume that only four classes are provided, which is 2 raised to the second power. In this case the four available amplitude classes consist of two for positive values of the signal and two for negative values. The four classes are transmitted by two pulses, two being the exponent. The four classes consist of (1) both pulses present, (2) only the first pulse present, (3) only the second pulse present, and (4) both pulses absent.

It has been found by experiment that satisfactory commercial telephone quality can be obtained by the use of a five-pulse system, which gives 2^5 or 32 amplitude classes. Of these, 16 are positive and 16 negative. Since the range of volume to be accommodated is about 60 db, this corresponds to approximately 2 db per amplitude class. Tests have shown this degree of quantizing to give a 33-db value of the signal-to-noise ratio; *i.e.*, some "quantizing noise" is present and prevents the signal-to-noise ratio from having a larger value.

In PCM the particular signal amplitude at a given moment is represented by the presence or absence of the individual pulses in a group of pulses. As an example, in a five-pulse system, if the first, third, and fifth pulses are present and the second and fourth pulses are absent, a particular amplitude value is indicated.

PCM is therefore characterized by the fact that all pulses transmitted have equal amplitude and equal duration; also (with the exception that certain pulses are omitted) they occur at equal intervals. Under these circumstances, noise is ineffective unless it is sufficiently strong to cause a pulse to appear where none was transmitted, or vice versa. This system is therefore highly resistant to noise and, under typical operating conditions, received signal strengths can be weaker than with other systems. At each repeater the restoring type of amplifier sends on the signal entirely free of noise; *i.e.*, noise is noncumulative as long as the incoming signal-to-noise ratio is above a modest threshold value.

In a typical system using five pulses or 32 amplitude values, the band width is somewhat less than five times that required for a pulse-amplitude-modulated system. This is the cost in band width to obtain the very large increase in ability to withstand noise.

On account of the complexity of the PCM equipment required to code the amplitudes at the transmitter and decode them at the receiver, the system is of special interest in connection with multiplex transmission, a time-division system being used in which the five pulses of one channel are transmitted, then the five pulses of the next channel, etc., the entire system operating with short pulses and short spacings so as to return to the first channel after an interval of $1/8,000$ sec or $125 \mu\text{sec}$. In a system having eight channels, this means that pulses are transmitted at the rate of $5 \times 8,000 \times 8$ or 320,000 per sec.

A special cathode-ray tube with a composite plate of rectangular shape has been developed² for the coding operation using a seven-pulse system and retaining the figure of 8,000 samples per second. This was demonstrated before the New York section of

¹ Grieg, D. D., Pulse Count Modulation System, *Tele-Tech*, September, 1947, pp. 48-52, 98; material also given under title Pulse-count Modulation, *Elec. Commun.*, September, 1947, pp. 287-296. Black, H. S., and J. O. Edson, PCM Equipment, *Elec. Eng.*, November, 1947, pp. 1123-1125. Clavier, A. G., P. F. Panter, and D. D. Grieg, PCM Distortion Analysis, *Elec. Eng.*, November, 1947, pp. 1110-1122. Coded Pulse Modulation Minimizes Noise, *Electronics*, December, 1947, pp. 126-131. Pulse Code Modulation, *Electronics*, August, 1947, pp. 136, 138, 174. Meacham, L. A., and E. Peterson, An Experimental Pulse Code Modulation System of Toll Quality, *Bell System Tech. J.*, January, 1948, pp. 1-43.

² Advanced Pulse Code Modulation System, *Tele-Tech*, November, 1947, pp. 31, 102-103.

the IRE in September, 1947, in a 96-channel multiplex system. This required pulses at the rate of over five million per second.

MOBILE RECEIVERS

Communication requirements of such special services as police,¹ taxis,² fire departments, public utilities, forestry services,³ etc., differ among themselves, but in general the receivers follow conventional engineering practice with certain special features, such as (1) fixed-frequency operation, (2) location of the r.f. generally in the v-h-f range, (3) squelch to keep the loud-speaker quiet between messages, and (4) convenient arrangements for changing between reception and transmission. Mobile radiotelephone apparatus often has facilities for the selective calling of particular cars and for tying into the general wire telephone system.

167. Police Service. The usual police service installed in recent years operates with phase modulation, equivalent to f.m. with a preemphasis of 6 db per octave. The carrier frequencies lie in one of the following bands: 30-40, 72-76, or 152-162 Mc. The deviation corresponding to 100 per cent modulation has been established as ± 15 kc for the two lower frequency bands and as ± 20 kc for the 152-162 Mc band. Two-frequency operation is the general rule, with one frequency for headquarters-to-cars transmission and the other for cars-to-headquarters transmission. Car-to-car feature is then obtainable by having a double-frequency transmitter or a double-frequency receiver in each car. In extensive systems,⁴ several receiving stations are necessary, the signal picked up by the various receiving stations being relayed to headquarters by wire line or over radio relay links using other frequencies. This constitutes a special-type large-scale diversity receiving system. In state-wide police systems, several transmitting stations are also usually required.

To avoid interference between police systems in various cities, the power transmitted by the headquarters equipment is generally limited, so that, with small antennas on the cars, high sensitivity in the receiver is required. As a typical figure, the squelch will open at signal inputs of 0.1 to 0.4 μ v. Quieting of 20 db is generally obtained with less than 1 μ v input.

A vertical antenna is generally used, mounted either at the rear of the car or in the center of the top. The latter position has the advantage of better omnidirectional characteristics. Mounting at the back has the disadvantage that the antenna is most sensitive toward the rear. The use of a loop antenna with an iron core has been proposed for police service.⁵

Police cars carrying radio receivers require the reduction of interference in much the same way as broadcast automobile receivers. Some difference exists because the frequency of operation is considerably higher.⁶

¹ Thomas, H. E., The Development of a Frequency-modulated Receiver for U-h-f Use, *RCA Rev.* October, 1941, pp. 222-233. This article gives a good discussion of design considerations.

² Freedman, Samuel, Two-Way Taxicab Radio Systems, *Communications*, December, 1947, pp. 10-13; February, 1948, pp. 8-11. Includes tabulation of technical data on equipment made by several firms.

³ Lawson, H. K., and L. M. Belleville, Mobile 30-40 Mc Receiver for the U.S. Forest Service, *Electronics*, January, 1942, pp. 22-25, 98-99. This describes a receiver for a-m operation having i.f. of 1.6 Mc. Simson, A. G., "The Communication Facilities of the U.S. Forest Service," paper presented at AIEE convention, August, 1941.

⁴ Noble, D. E., State-wide F-m Police System, *Electronics*, November, 1940, pp. 18-22; December, 1940, pp. 28-31, 66. Describes Connecticut state system with description of receivers in second installment. Doremus, John A., Planning a V-h-f Communications System, *Electronics*, September, 1943, pp. 96-101, 178, 180. Describes installation for state of Massachusetts. Cutting, Basil, The New Hampshire State Police Radio System, *FM & Television*, January, 1945, pp. 24-29, 87; Diversity Transmission on 75 Mc, *FM & Television*, February, 1947, pp. 26-27. Describes addition of relay station of Mt. Washington. Scavarda, C. J., Michigan's F-m State Police System, *Electronic Inds.*, February, 1944, pp. 90, et seq.

⁵ Polydoroff, W. J., Inductively Tuned Loop Circuits, *Radio*, April, 1946, pp. 21-22; May, 1946, pp. 20-22.

⁶ Price, H. G., Eliminating Car Noise in 28-Mc Mobile Reception, *QST*, May, 1947, pp. 37-40, 132.

168. Mobile Radiotelephone.¹ Receivers used in the mobile telephone service are very similar to police equipment except that a handset and a telephone bell are generally provided; also selective calling is usually employed so that the bell rings only in the desired vehicle. Connections may be made to the general wire telephone network of the country, and use of the equipment on the vehicle is practically like the use of an ordinary phone with the exception that a press-to-talk button must be held down while speaking and released while listening. A large number of cars have the same pair of transmitting and receiving frequencies. On account of this "party line" type of operation, it is necessary for the telephone user in the car to listen to see if the facilities are in use before making a call.

Frequency assignments in the band from 30 to 44 Mc have been made for highway mobile telephone, and assignments in the band from 152 to 162 Mc for urban mobile telephone. One frequency for station to vehicle and another for vehicle to station are assigned and together are said to constitute a "channel" for this service. The method of transmission is phase modulation. Channels were established every 60 kc with the expectation that only alternative channels would be used in a given locality. The great popularity of the mobile telephone service has forced consideration of the possibility of employing every channel in crowded localities.²

During the period from 1946 to 1948 there was a general trend in the mobile service toward more severe selectivity requirements. This included some reduction of the transmitted frequency deviation, increased receiver selectivity, and a reduction in spurious receiver responses. Published specifications for this service show this trend in part.³

169. Selective Ringing.⁴ For the purpose of ringing only the desired phone in the mobile service, Western Electric offers selectors with special features which enable them to operate at signal-to-noise ratios as low as 2 db. These selectors operate on alternate pulses of 600 and 1,500 cycles, the frequency changing from one to the other each time a pulse, such as produced by a telephone dial, is transmitted. Receiver filters are provided, one passing the 600-cycle signals and the other the 1,500-cycle signals. Each of these outputs is rectified, and each rectified output is applied to one winding of a polar relay. With this arrangement, the relay moves from one side to the other at each successive pulse. This action of the relay causes a code wheel to advance, this wheel having 23 notches and 23 holes. Pins are inserted in certain of these holes to identify the particular receiver and to determine what number will cause its bell to ring. Each telephone call consists of 5 digits not including 1, so selected as to add to a total of 23 digits. There are 2,030 such numbers, so that anyone of 2,030 cars operating on a given receiving frequency could be called. The first group of pulses advances the code wheels of all receivers tuned to the particular frequency. However, only those code wheels which have a pin in the hole corresponding to this number of digits will remain at this position; all other code wheels return to the start. The second number or group of pulses then carries the code wheels of the remaining receivers to a point corresponding to the sum of the first two digits. At this point, many of the code wheels will have no pin and so will return to zero. The third number likewise advances the remaining code wheels, after which some return to the start. In this way, after the 5 digits totaling 23, only one wheel is left; this wheel, reaching the twenty-third notch, rings the bell in the desired car. The number 1 is not used but is reserved as a clearing signal; *i.e.*, when it is transmitted, all code wheels return to zero as a reference condition.

¹ Rommes, H. I. and R. R. O'Connor, "General Mobile Telephone System," paper presented at AIEE convention, Chicago, November, 1947.

² Noble, D. E., Cavity Type Filters for Interference, *Tele-Tech*, January, 1948, pp. 34, *et seq.*

³ Martin, H. B., Frequency Modulation Mobile Radiotelephone Services, *RCA Rev.*, June, 1946, pp. 240-252. Butler, Frank E., Mobile Radiotelephone Service Links Nation, *Radio News*, May, 1947, pp. 45, *et seq.*

⁴ Anderson, C. N., and H. M. Pruden, Radiotelephone System for Harbor and Coastal Services *Proc. IRE*, April, 1939, pp. 245-253; see pp. 251-252. Cottrell, B. P., Selective Calling for Mobile Telephone Service, *FM & Television*, January, 1948, pp. 32-34.

A stepping switch made by C. P. Clare is used in the Hammarlund¹ selector. This has a total of 10 positions, so that the sum of digits of all call numbers in this system is 10. The digits are radiated as pulses of 6-kc modulation. If a 4-digit system is used, this provides 84 different calls; if 5-digit operation is used, there is a total of 126 different calls. This switch advances on each successive group of pulses, returning to zero at any point where the sum of digits so far does not equal the corresponding partial total from the number of the desired station. However, the Hammarlund system designates the particular station electrically by means of connections to certain of the positions on the switch. This can be done by a multiple connector plug inserted in the receiver. To change the number of a receiver therefore it is only necessary to remove the plug and put in another. Similarly whenever a set is removed from a vehicle for repair and a new receiver put in, it is only required that the old number plug be transferred to the new receiver.

The Hammarlund control can be operated, if desired, so as to have a lockout feature while a call is in progress. In this case a red light is flashed at all other vehicles, indicating that the circuit is busy, and the mobile equipment at all unwanted vehicles is inoperative, except that in an emergency such a locked-out mobile user can plug his headset into a special jack on his transmitter and send an urgent message, his receiver still not operating until headquarters so elects.

The reed type of relay² tuned to a particular audio note is the basis of selective calling in the system of the Federal Telephone and Radio Corporation. The particular a.f. amplitude-modulates a 7,000-cycle subcarrier of the transmitter, which then fully frequency-modulates the radiated carrier. In all, 810 different calls are available.

170. Fixed-station Receiver. Considerable refinement in the fixed-station receiver for mobile radiotelephone service is justified. One feature³ widely used is the codan, this term standing for "Carrier-Operated Device Anti-Noise." It allows the receiver to operate reliably under widely varying noise conditions without manual adjustment. The codan was first used in marine phone service from 2 to 22 megacycles and was subsequently applied to highway and urban service in the v-h-f region.

Another feature⁴ of the fixed-station receiver is the vogad, standing for "Voice-Operated Gain-Adjusting Device." This equipment is an audio a.v.c. and has the complexity necessary for operation without the presence of a carrier component. The vogad is useful in handling conversations involving various lengths of telephone line, speakers talking with various degrees of loudness and other variables.

RAILROAD RECEIVERS

Radio equipment is used in railroad service in the following ways: (1) two-way radiotelephone for dispatching use especially in freight yards and to a growing extent in main-line operation;⁵ (2) receiver operation for the reception of broadcast programs for the entertainment of passengers; and (3) passenger telephone service enabling calls to be made to and from any phone in the regular land telephone network.

171. Dispatching Service.⁶ Frequency assignments in the region from 158 to 162 Mc have been made for the operation of railroad dispatching telephone service. Ver-

¹ Kulansky, J. K., Selective Calling System, *Electronics*, June, 1946, pp. 96-99. Dorr, C. J., and H. M. West, Relay Control Circuits for Stepping Switches, *Electronics*, January, 1948, all even pages from 158 through 176. Sleeper, Milton B., Selective Calling in New York on 157 Mc, *FM & Television*, February, 1946, pp. 46-49.

² Hoffman, R. B., and E. W. Markow, Mobile F-m Communications Equipment for 30 to 44 Mc, *Communications*, June, 1947, pp. 28-29, 40-41; August, 1947, pp. 34-35. Ferrar, R. C., and J. M. Lanzer, Multi-channel Selective Calling, *Tele-Tech*, August, 1947, pp. 31-33, 97-99.

³ Fischer, H. B., "Remote-controlled Receiver for Radiotelephone Systems, *Proc. IRE*, April, 1939, pp. 264-269.

⁴ Wright, S. B., S. Doba, and A. C. Dickieson, A Vogad for Radiotelephone Circuits, *Proc. IRE*, April, 1939, pp. 254-257.

⁵ Freedman, Samuel, "Two-way Radio," Ziff-Davis Publishing Co., 1946; note Chap. 9, Induction Radio and Guided Carrier Systems, pp. 143-173, and Chap. 12, Two-way Radio for Railroads, pp. 229-318.

⁶ Sands, L. G., VHF Radio Equipment Speeds Up Railroad Operation, *Tele-Tech*, May, 1947, pp. 38-41, 110-111. Peterson, Joel, Railroads Plan Greater Use of Radio for Communications, *Tele-Tech*,

tical polarization and phase modulation are used. Published specifications¹ of the Rock Island Railroad are representative. A trackside portable² similar to the walkie-talkie of the Second World War is available for the use of men operating switches, handling the couplings between cars, doing repair work, etc. Typical characteristics for this service are a total weight of about 15 lb, superheterodyne circuit, and a vertical $\lambda/2$ antenna. The range is about 1 mile.

The v-h-f dispatching service is often called "space radio" to distinguish it from systems having an inductive characteristic, which are taken up in the next paragraph.

172. L-f Inductive Dispatching Service. Guided radio (*i.e.*, wired wireless) systems³ utilizing the rails and especially the normal wayside telegraph and telephone wires to propagate the waves offer the advantage of greater distance than obtainable with v-h-f radio operation. Distances as great as 50 to 100 miles between a wayside station and trains are typical.

These systems generally operate on frequencies between 50 and 200 kc and use frequency or phase modulation with a value of approximately unity for the modulation index.

In extended railroad yards an inductive system can be used for the remote control of an additional v-h-f transmitter-receiver. In one design⁴ a carrier of 189 kc was used with a deviation of ± 4 kc. The distance between the main and the satellite stations was 12 miles. The receiver was a t-r-f type requiring $10 \mu\text{v}$ to saturate, under which condition it delivered an output of 5 watts; the band width at 50 db was 20 kc and at 70 db was 26 kc.

173. Other Dispatching Services. A railroad dispatching system operating at 2,660 Mc⁵ has been reported to give good results. The receiver in this equipment employs a crystal mixer followed by a 7-Mc i-f amplifier having four regular stages, followed by two limiter stages. The i-f band width at 6 db is 300 kc. A special vertical antenna array of biconical parabolic reflectors was designed for this service.

At the opposite extreme in the frequency spectrum, is the unit train communication system of the Union Switch and Signal Company.⁶ This is an a-m system in which the transmitted signal is a single upper side band of a suppressed carrier such as 5,700 cycles. This signal at very low voltage and high current is fed conductively into the rails and picked up inductively from the rails. If paralleling telegraph or phone wires on the right of way are present, distances up to 10 miles can be obtained between trains and distances up to 100 miles between fixed stations and trains. The receiver in this system has two stages which amplify the received single side band and have a.v.c. A push-pull mixer, a peak limiter, and audio amplifiers are also provided.

Technique of identification (called IFF for Identification, *Friend* or *Foe*) developed during the Second World War has been proposed⁷ for collision prevention. A challenger is visualized in each locomotive and a transponder at the rear end of each train.

January, 1947, pp. 78-83, 150-151, 153. Halstead, W. S., FM for Railway Managerial Control, *FM & Television*, November, 1946, pp. 23-29, 57, 59. Sleeper, M. B., Status of Railroad Radio, *FM & Television*, June, 1946, pp. 19-20, 53. Courtney, Jeremiah, Railroad Radio from FCC to ICC, *Electronics*, June, 1946, pp. 92-94. Dahl, E. A., Rock Island Railroad Radio Tests, *Electronics*, May, 1945, pp. 96-102. Train Communications, *Elec. Eng.*, December, 1946, pp. 547-553. Gives data from Association of American Railroads.

¹ Railroad Radio Specifications, *FM & Television*, June, 1946, pp. 39-42, 59. Martin, David W., F-m Receiver Design for Rail Radio Service, *Communications*, August, 1947, pp. 14-17, 37.

² VHF Railroad Trackside Portable, *Tele-Tech*, June, 1947, pp. 43, 112.

³ Mesch, F. A., and L. H. Morse, FMI Railway Radio, *FM & Television*, September, 1947, pp. 30, *et seq.* The title refers to frequency-modulated inductive operation. Triem, W. R., Portable Inductive Radiophone, *Electronics*, December, 1947, pp. 93-95.

⁴ Halstead, W. S., Railroad F-M Satellite System, *Communications*, May, 1946, pp. 17-21, 54-55. Nygren, A. C., and W. G. Clinton, 161-Mc Satellite System for Railyards, *FM & Television*, April, 1946, pp. 38-43, 68.

⁵ Dahl, E. A., 2660-Mc Train Communication System, *Electronics*, January, 1946, pp. 118-122.

⁶ Grondahl, L. O., and P. M. Bossart, Train Communication, *Elec. Eng., Trans. Sec.*, July, 1943, pp. 493-500.

⁷ Jarvis, K. W., What Can Radio Do to Improve Railway Safety? *Proc. Nat. Electronics Conf.*, Chicago, 1946, pp. 637-643.

The range of the proposed system is 8 miles, and an indicator shows the distance to the nearest train ahead. Another system based on transponder technique is known as "RadIn" and has been suggested by the Sperry Gyroscope Company.

174. Broadcast Reception on Passenger Trains. Conventional home receivers have long been used in limited numbers on railroads for the reception of programs from standard m-f broadcast stations, the special provisions of such installations generally being limited to suitable antenna and power-supply arrangements. Recently special designs for this service have been made with very good a.v.c., low spurious responses, and other features.¹

MARINE, AMATEUR, AND POINT-TO-POINT RECEIVERS

Receivers designed for use in marine, amateur, and point-to-point service are often called "communication receivers"² to distinguish them from broadcast receivers. They differ markedly from broadcast sets in various respects. The general quality of construction is superior, particularly with regard to the use of longer lived components and to the frequent provision of increased selectivity for reception in the presence of heavy interference. An outdoor antenna rather than a loop antenna is used. Various special features and numerous controls are provided. The housing is generally a functional metal cabinet. In point-to-point receivers, special features and individual design for the particular service may be extensive.

175. Marine Receivers.³ The severe climatic conditions encountered in tropical and arctic waters make it necessary for marine receivers to utilize parts treated with special impregnating agents. In addition, a high degree of general reliability is required, as well as accessibility in case repairs are needed.

During the rapid construction of vessels in the Second World War, standard complete shipboard radio units⁴ were designed, which included a receiver for 15 to 650 kc, an emergency crystal-detector receiver for 500 kc requiring no power, and an autoalarm as a stand-by to receive distress signals on 500 kc.

Another type of marine equipment for ocean-going service is available to provide two-way telephone service at frequencies from 2 to 12 Mc, and especially for the range from 2 to 3 Mc. Except for the use of a.m. and the larger and heavier construction, equipment of this type is similar to that provided for the mobile telephone service. Equipments in this service often include preset tuning arrangements providing a choice of six to ten frequencies with facilities to change the tuning of both transmitter and receiver simultaneously. A push-to-talk button is provided, or a voice-operated relay may perform this function automatically. Selective ringing equipment is available as an additional feature.

176. Amateur Receivers.⁵ The manufactured sets offered for amateur service average about 12 tubes and provide various special features. A tuned r-f stage is generally included, to give improved image and i-f rejections and signal-to-noise ratio. A band-spread knob and dial are provided and permit band-spread operation at any point in the frequency range of the set, or in some designs in the amateur bands only. A switch to disable the a.v.c. and one to start a beat-frequency oscillator adapt the set for the reception of c-w telegraph signals.

The crowded conditions on the amateur bands make high selectivity a prime

¹ Klippel, K. L., and E. A. Dahl, Railway Entertainment System, *Electronics*, May, 1947, pp. 118-121. Curtis, John A., Passenger Entertainment System for Railroad Use, *Tele-Tech*, July, 1947, pp. 34-37, 96-97.

² Gridale, G. L., and R. B. Armstrong, Tendencies in the Design of the Communication Type of Receiver, *J. IEE (London)*, Part III, September, 1946, pp. 365-384.

³ Nilson, A. R., and J. L. Hornung, "Practical Radio Communication," 2d ed., McGraw-Hill, 1943. See Chap. 14, pp. 722-753.

⁴ Standardized Marine Radio Unit, *Electronics*, January, 1942, pp. 36-39. Lewis, E. F., Packaged Marine Radio, *Electronic Inds.*, October, 1943, pp. 88-89, 152, 154.

⁵ "Radio Amateur's Handbook," American Radio Relay League, West Hartford, Conn.; see chapter entitled "Receiver Principles and Design." Bray, W. J., and W. R. H. Lowry, The Testing of Communication-type Radio Receivers, *J. IEE (London)*, Part IIIA, No. 12, 1947, pp. 313-327. Villard, O. G., Jr., Simplified Single-sideband Reception, *Electronics*, May, 1948, pp. 82-85.

requirement. For this reason, the amateur type of receiver includes in the i-f amplifier a filter having a quartz crystal as one element. The usual i.f. for an amateur receiver is 455 kc, so that this is the natural frequency of the crystal. The very sharp mechanical resonance of this crystal permits band widths as small as 200 cycles to be obtained, thus permitting the reception of telegraph signals through very severe interference. This very sharp selectivity has led to the name "single-signal" receiver.¹ A selectivity control is provided to permit various greater band widths so that phone messages may be received with the best quality permitted by the prevailing conditions.

A noise limiter is generally found in the amateur receiver and is useful for reducing ignition interference.

177. Point-to-point Receivers. The particular requirements often existing in point-to-point service have led to many special receiver designs, some of which are very elaborate. The use of diversity antenna systems is found in this service.

An outstanding case of point-to-point receiver design is the transoceanic radiotelephone equipment used by the Bell System.² The receiver in this equipment operates on any frequency from 4 to 22 Mc and provides as many as three one-way voice channels in a radio band having a total width of 12 kc. Each channel is of commercial telephone quality with a width of about 3,000 cycles.

These channels are obtained as follows: (1) a side band somewhat removed from the carrier on one side furnishes one channel; (2) a similar displaced side band on the other side of the carrier furnishes the second channel; and (3) a third channel is "straddled" across the carrier.

The two outer side bands are moved away from the carrier to the extent that no part of either side band is within about 2,200 cycles of the carrier. All three channels utilize the standard commercial privacy equipment which divides the voice range into five equal frequency bands and interchanges these, with or without inversion, making a change in the arrangements every 20 sec. In the case of the third channel, which is straddled across the carrier, two of these privacy segments are on one side of the carrier, and the remaining three are on the other side.

To receive and separate a triple transmission of this kind is the function of a receiver having about 58 tubes and various special features. The design is a double super-heterodyne with i.f.s of 2,900 and 100 kc. The transmission employs carrier of reduced amplitude. The receiver can therefore be operated to produce and use reconditioned carrier obtained from the incoming signal, or it can produce and use a standard 100-kc frequency as a resupplied carrier. Crystal filters operating in the neighborhood of 100 kc provide sufficient selectivity for separation of the carrier and the various desired side bands. The receiver includes a mechanical type of a.f.c. which operates on the first heterodyne oscillator so as to maintain the frequency of the 100-kc carrier in the second i-f amplifier.

Point-to-point practice in the v-h-f range may utilize f.m., as in the case of the installation (42 to 50 Mc) to serve the Queen Charlotte Islands of British Columbia.³ Frequencies from 35 to 80 Mc have been used in point-to-point practice by the British.⁴

178. U-h-f and S-h-f Relay Receivers. A relay system operating at 2,000 Mc has been installed by General Electric between New York and Schenectady.⁵ This system operates with f.m. and a maximum deviation of ± 7 Mc. The receiver resembles those in radar practice and has a 1N21B crystal mixer, an SRL-7A klystron super-heterodyne oscillator, and an i-f amplifier operating at 100 Mc. The i-f amplifier

¹ Single Signal C. W. Reception and Crystal Filters, *QST*, March, 1946, pp. 59-61.

² In this connection, see A. H. Mumford Long-distance Point-to-point Communication, *J. IEE (London)*, Part III A, 1947, pp. 23-43, W. J. Bray, H. G. Lillierap, and W. R. H. Lowry, The Design of Transmitter Drives and Receivers for Single-sideband Systems, *J. IEE (London)*, Part IIIA, 1947, pp. 298-312.

³ Knox, J. B., and C. H. Brereton, A Multi-channel VHF Radio Communications System, *RCA Rev.*, June, 1946, pp. 179-198.

⁴ Hollinghurst, F., and C. W. Sowton, Résumé of V.H.F. Point-to-point Communication, *J. IEE (London)*, Part IIIA, 1947, pp. 115-131.

⁵ Deerhake, F. M., 2000-Mc Television Program Chain, *Electronics*, February, 1948, pp. 94-97.

includes three triplets, each consisting of three stages. Each triplet is stagger-tuned at 92, 100, and 108 Mc. In this way a band width of 16 Mc is obtained. These three triplets are followed by a two-stage limiter, which is followed by a discriminator for detection of the video signals, the purpose of the relay system being to transmit television programs.

A 4,000-Mc relay system for television and multiplex telephone service has been opened by the Bell System between New York and Boston. The receivers in this system also resemble radar practice, but are notable in the use of the Western Electric 404-A pentode, a tube with about twice the transconductance of the 6AK5 and less noise, while the input and output capacitances remain about the same.

Point-to-point equipment for television pickup service at 7,000 Mc has been designed by RCA.¹ This operates with f.m. The receiver is divided into two parts, the first of which is mounted directly behind the parabolic "dish" which constitutes the receiving antenna. This part of the receiver includes a crystal mixer, a 2K26 klystron oscillator, and four i-f stages operating at 120 Mc. The band width is 20 Mc. A.f.c. with a d-c amplifier is provided, acting on the reflex electrode of the klystron. The second unit of the receiver furnishes additional i-f amplification, as well as discriminators for the signal circuit and for the a-f-c operation. The i.f. corresponding to the peaks of the television synchronizing signal is held by means of the a.f.c. to an accuracy of ± 1 Mc regardless of the picture content.

MILITARY AND NAVAL RECEIVERS

Special attention to reliability despite severe conditions of temperature, humidity, mechanical shock, and vibration is necessary in the design of military and naval² radio equipment. In addition, in case the apparatus is for use in aircraft, it must operate at greatly reduced air pressure or be housed in an airtight case raising problems of heat dissipation. Numerous specifications have been prepared giving details of these various requirements, and compliance with such specifications is required according to the contracts under which equipment is purchased by the Government from manufacturers. Special attention is necessary on the part of the design engineer to ascertain definitely what specifications must be observed and to work out satisfactory designs.

During the Second World War many new names and abbreviations relating to military and naval radio, radar, aids to navigation, countermeasures, and related topics came into use. Glossaries of these are available in the literature.³

179. Typical Military Requirements. Selected provisions from various specifications will serve to indicate some of the special features required in radio equipment for the defense services. The precise requirements in each case depend on the specifications listed in the contract applicable to the particular equipment, the choice of the particular specifications being made by the Government authorities on the basis of the character and purpose of the equipment and the state of military and radio knowledge at the time.

Temperature. A typical requirement is that operation must be satisfactory for extended periods at any point from -55°C (-67°F) to $+55^{\circ}\text{C}$ ($+131^{\circ}\text{F}$). Operation at higher temperatures may be required in some cases. Resistance to high humidity is generally required, such as exposure to 100 per cent humidity at a temperature of 50°C (122°F).

Pressure. For air-borne equipment, operation at a pressure of 3.4 in. Hg, corresponding to an altitude of 50,000 ft, is a representative requirement.

¹ Poch, W. C., and J. P. Taylor, *Microwave Equipment for Television Relay Service*, *Broadcast News*, October, 1946, pp. 20-27.

² Colton, R. B., *Army Ground Communication Equipment*, *Elec. Eng.*, May, 1945, pp. 173-179. Kinert, J. O., *Naval Wartime Communication Problems*, *Proc. IRE*, April, 1946, pp. 193W-195W. Davis, T. M., and E. Toth, *Design of Communication Receivers for the Naval Service with Particular Consideration to the V-h-f and U-h-f Ranges*, *Proc. IRE*, February, 1947, pp. 201-207. Read, C. T., *Modifying Radio Equipment for Military Applications*, *Electronics*, February, 1944, pp. 98-101.

³ DeVore, Charles, *Wartime Words and Their Meanings*, *Electronics*, June 15, 1947, pp. 108-113. Roberts, Arthur, "Radar Beacons," pp. 475-477, McGraw-Hill, 1947.

Vibration. Requirements may include the range from 10 to 55 cps with the equipment being tested at a total excursion of 0.06 in. The design of flexible mountings for resisting vibration and shock is quite important, and it was found¹ during the war that suitable designs would attenuate vibration in the region of 10 to 30 cycles to a value of less than 1 per cent, and vibration of 30 to 200 cycles to a value of less than 0.1 per cent.

Shock. Resistance to shock is necessary so that the equipment shall not be readily disabled by the rough landing of aircraft, enemy bombs bursting on the decks of vessels, or other causes. Shock corresponding to an acceleration of 25 g is sometimes stated as a value which must be successfully withstood. Another specification, in dealing with equipment weighing over 250 lb, requires that it be mounted on a table weighing about 2 tons, and that the table then be struck underneath with a 3,000-lb hammer moving at a velocity corresponding to a fall through a height of 9 to 66 in., the value of this equivalent fall varying with the weight and class of the equipment.

Components. Inductors and transformers having laminated-iron cores are subject to severe requirements and must pass tests for temperature rise of the windings in operation, as well as exposure to extreme external temperatures and immersion in successive baths of hot and cold water. The potting compound must not crack at -55°C (-67°) nor flow at $+85^{\circ}\text{C}$ ($+185^{\circ}\text{F}$).

Variable capacitors must have a nominal clearance between fixed and moving plates of not less than 0.012 in. and an actual clearance not less than 0.008 in.

Electrolytic capacitors are normally usable only if approved by the Government after a showing by the contractor. In case their use is approved, the plug-in type is preferred but must be provided with means for firmly holding the capacitor in its socket.

Resistance to corrosion must include the ability to withstand exposure to a salt sea atmosphere. In the case of steel, phosphate treatment is generally acceptable. Another acceptable finish on steel is a cadmium plating of 0.0005 in. thickness followed by either a dichromate treatment or lacquering. Aluminum can be treated by (1) anodizing, (2) a caustic dip and lacquer, or (3) sandblast and lacquer.

Fungus. The war in the Pacific demonstrated that resistance to the growth of fungus² is one of the most important requirements for military equipment to be operated in the tropics. It is necessary to avoid the use of fungus-nutrient material as far as possible and, where this is impossible, to treat with a fungicide. The thickness of the fungicide spray after drying should be 0.002 in.

Receiver radiation must be less than 400 μW according to the usual Naval requirement.³

The soldering of wires must normally follow a secure mechanical attachment, so that the soldering is not depended on for mechanical strength. The use of excessive solder is not acceptable. Not over three wires may be attached to one lug, but multiple lugs may be used to whatever extent necessary.

180. Experience in Severe Cold. Expeditions and tests in cold climates⁴ have shown that radio, radar, and other electronic equipment are useful, but that the available models were subject to certain limitations. At one location in the Antarctic, situated on a thickness of several hundred feet of ice, local interference due to ignition and commutator noise was severe because the thickness of ice prevented the use of a ground connection. Another difficulty encountered was the unloading and moving into position of heavy diesel-engine generators for GCA systems. Other tests in severe

¹ Berkner, L. V., Naval Airborne Radar, *Proc. IRE*, September, 1946, pp. 671-706; note p. 692 regarding vibration and shock.

² Proskauer, R., and H. E. Smith, Fungus and Moisture Protection, *Electronics*, May, 1945, pp. 119-123. Healy, C. P., and J. C. Niven, Mould and Humidity in Radio and Signals Equipment, *Proc. IRE*, May, 1945, pp. 300-306. Proskauer, R., Fungus-proofing Procedure, *Electronics*, June, 1944, pp. 92-93, 224, 229, 232.

³ MacDonald, G. J., and D. A. Thorn, Radiation from Receivers, *J. IEE (London)*, Part IIIA, No. 12, 1947, pp. 427-436.

⁴ Bailey, H. C., Electronics in the Antarctic, *Electronics*, August, 1947, pp. 82-88. "Military Communication Equipment Observed at Winter Tests," report based on data of M. R. Johnson, S. W. Jucius, and R. B. Zechiel, *Elec. Eng.*, August, 1947, pp. 781-783.

weather showed that the handicap of extreme cold greatly reduces the efficiency with which personnel are able to install, calibrate, operate, maintain, and repair electronic apparatus.

181. Receivers for Identification and Beacon Service. Identification and beacon systems¹ include (1) an interrogator-responder (also called an "I-R," or "challenger") which initiates the operation, and (2) a transponder (also called a "replier" or "responder"), which automatically answers the interrogation. An I-R is a transmitter-receiver, and a transponder is a receiver-transmitter. Identification systems operate as accessories to radar. After the radar has shown the presence of a distant plane or vessel, the identification equipment enables the friendly character of the distant craft to be checked. All friendly craft must therefore carry transponders. In addition, craft that are expected to make attacks on the enemy must be equipped with interrogator-responders so as to make a check of identity before proceeding with an attack. Normally therefore interrogator-responders are furnished in one or more of the craft in any attacking group.

The receivers of I-Rs are superheterodynes very similar to those in radar practice. The receivers of transponders have generally been superregenerators with AGS.

A radar beacon system² is much like an identification system. The moving craft sends out the interrogation and the beacon returns the reply; i.e., the I-R is on the craft and the transponder is at the known point where it acts as the beacon. The craft learns its location by (1) knowing the identity of the beacon from the special coding of the reply, (2) observing the location of the reply on the radar scope, and (3) taking advantage of the fact that, starting with the known location of the beacon, the reverse bearing and the same distance determine the point on the map where the craft is located.

182. Countermeasure Receivers.³ To reduce the value of radio and radar to the enemy, it may be necessary to (1) ascertain the frequency of the transmission and (2) radiate an interfering signal of sufficient strength and suitable modulation to prevent the enemy from realizing any benefit from his efforts. Special search receivers are required for the first of these functions. Another aspect of countermeasure-receiving technique is the use of two or more direction finders so as to ascertain the location of an enemy transmitter.

Search receivers for finding enemy radar signals⁴ were developed during the Second World War, the designs being superheterodynes covering the range from 25 to 10,000 Mc. Butterfly tuners and acorn tubes were used up to about 1,000 Mc with an i-f amplifier of 3 to 4 Mc band width centered at 30 Mc. Image rejection was over 40 db at the lower frequency and over 10 db at 1,000 Mc. From 1,000 to 10,000 Mc, concentric-line oscillators with reflex klystrons were used. A 200-Mc i-f amplifier having a band width of 20 Mc at the peak and 24.5 Mc at 6 db, used with a two-cavity preselector, gave image rejections of 40 to 60 db. Reflector voltages were about 300 at the lower end of the band, increasing to 1,250 at the upper end. These equipments were air-borne and, with associated jammers, greatly reduced the effectiveness of German anti-aircraft radar fire control.

Wartime countermeasure research and development centered at the Harvard Radio Research Laboratory, and the general engineering findings of this work are available in the literature.⁵

¹ Stuart, Lois E., Identification, Friend or Foe—Radar's Sixth Sense, *Tele-Tech*, January, 1947, pp. 60-67. Wood, K. A., 200-Mc/s Radar Interrogator-Beacon Systems, *J. IEE (London)*, Part IIIA, No. 2, 1946, pp. 481-495.

² Roberts, *op. cit.*

³ A nontechnical account of radar countermeasures is given in James Phinney Baxter, III, "Scientists against Time," Little, Brown, 1947; see pp. 158-169.

⁴ Radar Countermeasures, *Electronics*, January, 1946, pp. 92-97. Sinclair, D. B., Search Receivers for Radar Countermeasures, *Gen. Radio Experimenter*, March, 1947, pp. 1-5. Villard, O. G., Jr., Radar Countermeasures: The Science of Immobilizing Enemy Radar, *Proc. Radio Club of America*, March, 1946, pp. 7-15.

⁵ Staff of Radio Research Laboratory, Harvard University, "Very High Frequency Techniques," McGraw-Hill, 1947.

Panoramic equipment is of special interest in search receivers and is described below.

183. Antijamming Features in Receivers. As a counter-countermeasure it is essential to design military receivers to be as immune as possible to enemy jamming. The first antijamming feature is naturally against unmodulated c-w interference. Such immunity is obtainable to a good degree in a radar receiver by changing from a d-c coupling to a capacitance-resistance type between the detector and the first video stage. This prevents the first video stage being biased beyond cutoff by the rectified jamming signal.¹

Certain types of modulation on the interfering signal can be much reduced by switching in a short-time-constant coupling (a differentiator) between the detector and the first video stage. For example, the time constant can be reduced to only $1 \mu\text{sec}$, assuming that the desired signal consists of short pulses.

If the interference overloads the i-f amplifier at the times of high a.m. of the interference, the condition can be improved by fast a.v.c., also called "instantaneous automatic gain control" (IAGC) or "back bias."² Such controls can be applied to one or more of the last i-f amplifier stages.

The IAGC circuit is also useful in reducing "clutter" on the radar scope, such as due to echoes from land, sea, or clouds. Another means for reducing clutter is sensitivity-time-control (also called "STC," and "GTC"), which reduces the receiver sensitivity immediately after each transmitter pulse and then rapidly increases the sensitivity as more distant echoes arrive.

The a.f.c. of a radar receiver would be subject to jamming by c-w interference if a separate a-f-c crystal mixer were not provided.³ By including such a separate crystal, it can be arranged to deliver the difference between the transmitter frequency and the local oscillator, so that the a.f.c. is unaffected by returning echoes or incoming interference.

DIRECTION FINDERS

The directive properties of loops and other types of antennas make possible the design of direction finders⁴ for the various portions of the radio spectrum. A direction finder is a radio receiver with a rotatable antenna or other means for determining the direction of arrival of a wave from a distant transmitter. Many instruments indicate first the direction of the infinite straight line passing through the transmitter and the direction finder, after which some further manipulation by the user is required to ascertain which end of the line is toward the transmitter, *i.e.*, to remove the 180-deg ambiguity. A more advanced type, called an "automatic direction finder" or ADF, recently introduced, includes a seeking feature so that, after a station is tuned in, the needle of a dial shows the direction without ambiguity. Very recent types show the direction on a cathode-ray tube and are constantly sweeping the horizon so that very brief transmissions will be seen.

Direction finders with scope indicators operating in and below the v-l-f region have been found useful for locating storms over ocean areas.⁵ One direction finder for this service has a tuning range from 3.6 to 17.5 kc.

184. Typical Marine Direction Finders.⁶ The law requires certain classes of vessels to be equipped with direction finders for the range from 285 to 515 kc. This is

¹ Ridenour, L. N., "Radar System Engineering," McGraw-Hill, 1947; see pp. 457-460.

² Schneider, E. G., Radar, *Proc. IRE*, April, 1946, pp. 528-578; see pp. 554-555. See also Ridenour, *loc. cit.*

³ Straus, H. A., *et al.*, The MPG-1 Radar, *Electronics*, January, 1946, pp. 110-117; see p. 114.

⁴ Keen, R., "Wireless Direction Finding," 4th ed., Iliffe & Sons, London, 1947. Bond, D. S., "Radio Direction Finders," McGraw-Hill, 1944. *J. IEE (London)*, Part III, March, 1947, contains several articles on direction finders. *J. IEE (London)*, Part IIIA, No. 11, 1947. Two papers and a number of summaries on direction finders are given on pp. 132-168.

⁵ Static D-f Reveals Storm Location, *Electronics*, February, 1946, pp. 212, 214. More on Spheric Storm Detector, *Electronics*, April, 1946, pp. 224, 226, 228. Adcock, F., and C. Clarke, The Location of Thunderstorms by Radio Direction-finding, *J. IEE (London)*, Part III, March, 1947, pp. 118-125, with discussion starting on p. 133. Kessler, W. J., and H. L. Knowles, Direction Finder for Locating Storms, *Electronics*, May, 1948, pp. 106-110.

⁶ Nilson and Hornung, *op. cit.*, 2d ed., Chap. 16, pp. 780-817.

sufficient to operate on the marine beacons at 285 to 315 kc, the ship-to-ship compass frequency of 375 kc, and the international distress frequency of 500 kc.

The typical marine direction finder has a shielded loop antenna, a simple antenna for sense (180 deg-ambiguity) determination, an r-f stage for each half of the loop (the center of the loop being grounded), a highly selective receiver, and headphones. The receiver is generally a superheterodyne of 6 or 8 tubes, and 175 kc is a typical i.f. Sensitivity values are typically 80 to 100 db below 1 volt per m (100 to 10 μ v per m) for 6 mw out at 300 kc. The sharp selectivity can be seen from the representative 300-kc band widths of 6 kc at 20 db, 10 kc at 40 db, and 14 kc at 60 db.

185. Automatic Direction Finders.¹ In aircraft navigation, bearings with a direction finder are wanted on a rapid and continuously indicating basis.² This need is satisfied by the ADF which has come to be a standard type of instrument. These equipments furnish nonambiguous readings. A servo system automatically keeps the loop and an indicating needle oriented for the desired station as the plane proceeds. At least one dual type is available, in which an additional antenna and receiver are provided to operate a second needle of the indicator, so that the bearings of two stations are continuously shown. The frequency coverage of ADFs is usually from 150 to 1,700 kc, and the i.f. about 112 kc. The sensitivity is about 86 db below 1 volt per m (50 μ v per m). The loop antennas for ADFs are usually shielded and also housed in a specially coated radome to reduce precipitation static. A loop core of suitable iron has been proposed and is stated to improve the performance or reduce the size of the loop.

186. H-f Direction Finders.³ During the Second World War there was intensive development of direction finders for quickly taking bearings on German submarines which transmitted messages on frequencies in the h-f spectrum. Various models accommodated the range from 1.5 to 30 Mc, and included scope presentation of the polar type, the DF station being represented at the center and an indication in any direction showing instantly the bearing of the distant transmitter.

These equipments have two fixed antenna systems oriented at right angles to each other and connected to the two primaries of a special transformer called a goniometer. An accurate right-angle relation between the two primaries is established in the construction of the goniometer. The secondary winding rotates in the field of the two primaries and therefore picks up signal in the same way as a loop antenna rotating in space. A motor spins the secondary at high speed and also controls the scanning of the scope.

- These equipments are of peacetime value for the location of planes or vessels in distress, so that aid can be dispatched, especially over great ocean areas.

AUTOMATIC ALARM RECEIVERS

An interesting type of receiver⁴ is used on vessels and at shore stations to receive automatically the international distress signal at 500 kc and sound alarm bells. The international signal for this purpose, established at the Madrid Telecommunications Conference of 1932, is a succession of dashes, each lasting 4 sec, separated by spaces of 1 sec. American designs operate upon the receipt of four such dashes, and European designs upon the receipt of three.

187. Auto Alarm Specifications. The FCC requires these receivers to operate at any frequency between 487.5 and 512.5 kc. When tested with a dummy antenna consisting of 500 μ mf, 20 μ h, and 5 ohms in series, the alarm must operate with a 30 per

¹ Scott, Allen, New Four-Band ADF, *Air Transport*, April, 1946, pp. 67-68, 71. Describes Model AN/ARN-6. Bond, D. S., Visual Direction Finders, *Electronics*, November, 1943, pp. 140-143; December, 1943, pp. 140-146, 324-325; January, 1944, pp. 144, *et seq.*

² Levy, G. F., Loop Antennas for Aircraft, *Proc. IRE*, February, 1943, pp. 56-66; July, 1943, p. 384.

³ Ross, W., Fundamental Problems in Radio Direction Finding at High Frequencies (3-30 Mc/s), *J. IEE (London)*, Part IIIA, No. 11, 1947, pp. 154-168. H-F Direction Finder, *Electronics*, March, 1946, pp. 304, 306, 308. Hansel, P. C., Instant-reading Direction Finder, *Electronics*, April, 1948, pp. 88-91.

⁴ Automatic SOS Alarms, *Electronics*, April, 1937, pp. 20-23.

cent modulated signal of 500 μv . Assuming the antenna to have an effective height of 5 m, this corresponds to a signal strength of 100 μv per m. A warning must be given for various defective conditions, such as a burned-out filament. Means for checking with a local test signal must be provided. The equipment must withstand various conditions of temperature, vibration, etc.

The RCA auto alarm¹ is a nine-tube superhet of the infradyne type, the i.f. being 1,100 kc and the oscillator frequency 1,600 kc. The circuit includes a 6A8 converter, two 6K7 i-f stages, a 6H6 detector, and five 1611 tubes (6F6 with closer tolerance on transconductance). Four 1611 tubes operate various relays to distinguish the distress call from miscellaneous interference and atmospherics, and the remaining 1611 tube is a voltage regulator. The sensitivity at the maximum position of the sensitivity control is 200 μv or 40 μv per m on the basis of a 5-m effective antenna height.

The Mackay auto alarm is a five-tube t-r-f design with a 6D6 r-f stage, a 76 detector, two 76 a-f stages, and an 89 connected as a triode detector to produce d.c. from the audio signals. Fixed over-coupled circuits are used between the antenna and the r-f stage, and also between the r-f stage and the detector. The necessary width to accept signals from 487.5 to 512.5 kc is thus obtained with adequate rejection outside this band. A buzzer, provided as a source of test signal, delivers about 2,000 μv to the antenna. The selector, which distinguishes the distress call from interference and static, consists of three motor-driven timing elements and a group of counting relays.

RECEIVERS FOR USE IN AIRCRAFT

The design of receivers² for use in aircraft requires special attention to obtain light weight, suitable housing, resistance to vibration, adaptability to aircraft antennas, operation with available power sources, and resistance to interference conditions.³ In addition, remote control is often required.

As part of the research work done during the Second World War, a careful comparison of a.m. against f.m. for aeronautic radiotelephone service was made at Harvard University for the Army Air Forces.⁴ This work included comparisons using two receivers which were similar except for the limiter and detector. Comparisons were made under various conditions of r-f signal strength, random-noise interference, pulse interference, and with the receivers slightly off tune. The general conclusion was favorable to f.m.

188. Shielding and Bonding.⁵ The ignition system of an airplane motor produces vigorous electrical effects characterized by steep wave forms which contain frequency components throughout a very wide range. These would induce an overwhelming amount of interference into any radio antenna on the plane unless they were confined by thorough shielding of the entire ignition system. In addition, shielding or filtering, or both, of most of the other wiring on the plane is required to avoid radio interference. It is especially necessary to filter leads which are connected to vibrating regulator contacts or to d-c motors. Small r-f filters are advisable in the receiver in series with each power-supply wire from outside, and also in series with external leads for audio connections, remote control, and other purposes.

Another essential element for radio reception in planes is that the various metal parts of the craft, especially moving parts, and its accessories must be firmly bonded together, as by the use of heavy copper braid.

¹ Byrnes, I. F., and H. B. Martin, Automatic Alarm, *RCA Rev.*, January, 1937, pp. 49-64.

² Nichols, R. G., Pilots' Radio Manual, *Civil Aeronautics Bull.*, Government Printing Office, Washington, September, 1940. Sandretto, P. C., "Principles of Aeronautical Radio Engineering," McGraw-Hill, 1942. Roberts, H. W., "Aviation Radio," Morrow, 1945. Morgan, H. K., "Aircraft Radio and Electrical Equipment," Pitman, 1939.

³ Trumbull, A. F., Aircraft Radio Design, *Electronics*, November, 1943, pp. 98-101, 220, 222.

⁴ Comparing FM with AM for Aircraft Communications, *Tele-Tech*, April, 1947, pp. 52-56, 110.

⁵ Frick, C. W., and S. W. Zimmerman, Radio-noise Filters Applied to Aircraft, *Elec. Eng., Trans. Sec.*, September, 1943, pp. 590-595. Foulon, F., Radio Noise Elimination in All-metal Aircraft, *Elec. Eng., Trans. Sec.*, 1943, pp. 877-891. Weinstein, G., et al., Radio-noise Elimination in Military Aircraft, *Elec. Eng., Trans. Sec.*, November, 1944, pp. 793-795. Sonbergh, G., VHF Ignition Noise, *Electronic Inds.*, November, 1944, p. 94-97, 200.

189. Precipitation Static. When flying is undertaken through snow, rain, ice crystals, or dust, especially at the higher speeds, precipitation static is encountered. This is due chiefly to neutral particles striking the plane, where they are ionized by friction, furnish a high electrostatic potential to the plane, and thus cause the plane to discharge into the surrounding air and produce radio interference. The potential of the plane may rise to a quarter of a million volts or more in severe cases. The discharge which occurs is corona, sometimes known as "Saint Elmo's fire." The effect on radio reception is worst if such discharge takes place from the antenna itself, but may still be severe if the discharge occurs only from the plane itself.

An extensive investigation of precipitation static was conducted by the Naval Research Laboratory, the Aircraft Radio Laboratory, and the University of Minnesota¹ using d-c potentials as high as 1,200,000 volts. The results of this and other work have been embodied in a report of the Radio Technical Commission for Aeronautics.²

Discussion pro and con of the use of very sharp discharge points is available in the literature.³

190. Loop Antennas for Aircraft Reception.⁴ The directional properties of the loop antenna and its independence of the electric field when shielded make it a valuable type of aircraft antenna for the lower frequency ranges. Sizes up to a maximum diameter of about 12 in. have been used. Shielding is always provided, located either near the wires or as part of a streamlined housing. In case the shielding is close to the wires of the loop, the radome should be slightly conductive. Loops of 4 to 20 turns are generally used and are called "low-impedance" in distinction to designs with a greater number of turns which have been used to a limited extent. The design of the loop can be solenoidal (all turns have the same radius), or pancake, meaning that there is a gradually changing radius with all the turns in the same plane. Composite designs with both solenoidal and pancake features can be used.

Aircraft loops with iron cores have been extensively used in Europe. This design is reported to offer the advantages of smaller size and suitability for placing nearer the metal fuselage of the plane.

191. Use of 28-volt Plate Supply.⁵ The 24-volt airplane battery furnishes a supply averaging about 28 volts while being charged. This is sufficient potential for the plate supply of certain tubes specially designed for the purpose. Care to avoid difficulties due to improper grid current is necessary in the operation of tubes at this low plate voltage.

The 28D7 can be operated as a low-power high-voltage supply by the provision of an oscillator circuit and a rectifier. Outputs up to 725 mw are obtainable at potentials of 50 to 250 volts, and smaller outputs are obtainable up to 500 or 600 volts. The frequency of oscillation can be located between 4 and 10 Mc. As the rectifier, the 14A4 tube can be used up to 300 volts, and the 28Z5 for higher voltages. By this use of the 28D7, limited amounts of high-voltage current can be obtained while still utilizing only a 28-volt source of power.

¹ This investigation, called the Army-Navy Precipitation Static Project, was reported in six technical papers in the *Proc. IRE*, April, 1946, pp. 156P-177P; May, 1946, pp. 234-254. See also Dana, H. J., Corona Discharge at High Altitude and Its Control to Reduce Interference, *Proc. Nat. Electronics Conf.*, Chicago, 1947, pp. 40-46.

² Reduction of Precipitation Static Interference in Aircraft, Paper 68-47/DO-5 of R.T.C.A., May 19, 1947.

³ Beach, R., Electrostatic Ills and Cures of Aircraft, *Elec. Eng.*, April, 1947, pp. 325-334; May, 1947, pp. 453-462. Beach, R., What of Air Safety? *Elec. Eng.*, May, 1948, p. 423. Wilcox, J. H., discussion of Beach paper, *Elec. Eng.*, October, 1947, pp. 1044-1046. Hall, W. C., Electrostatic Dischargers for Aircraft, *J. Applied Phys.*, August, 1947, pp. 759-765.

⁴ Levy, G. F., Loop Antennas for Aircraft, *Proc. IRE*, February, 1943, pp. 56-66, with correction in July, 1943, issue, p. 384. Burgess, R. E., Iron-cored Loop Receiving Aerial, *Wireless Engr.*, June, 1946, pp. 172-178. Polydoroff, W. J., Iron-core Loop Antennas, *Electronic Inds.*, November, 1943, pp. 84-85, *et seq.*

⁵ Tube Characteristics with 28-volt Plate Supplies, *Electronics*, December, 1947, pp. 190, 192, 194. Hammond, C. R., E. Kohler, and W. J. Lattin, 28-volt Operation of Receiving Tubes, *Electronics*, August, 1944, pp. 116-119, 379.

PANORAMIC RECEIVERS

A panoramic receiver¹ has an oscilloscope tube and suitable circuits to show the intensity of the signals available at the antenna over a frequency spectrum extending considerably above and below the frequency being received. This feature is obtained by means of an auxiliary channel which has amplifiers, a mixer, a swept oscillator, and the scope. The coupling into the auxiliary channel is from the modulator of the receiver and is made through a broad-band circuit so that signals of higher and lower frequency than the one being received will come through and appear with uniform strength on the scope.

192. Panoramic Details. The mixer of the auxiliary channel produces a new i.f. which then goes into a sharply tuned amplifier, the output of which is applied to the vertical plates of the scope. A sweep oscillator produces a saw-tooth output which is applied to the horizontal plates of the scope for scanning. The saw-tooth wave also drives a reactance tube so as to vary the frequency for which the auxiliary channel is sensitive. In this way the spectrum of interest is scanned, the frequency usually being 30 cps so as to produce a steady pattern on the scope. The horizontal axis shows the frequency at any point. Where a signal exists, an upward deflection or "pip" will be seen. If, for example, five transmissions in the frequency range of interest are taking place, five corresponding pips will appear. Furthermore, from the shape and time variation of each pip, the experienced observer can ascertain various information regarding the type of modulation and general character of the signal.

The panoramic feature is useful in communications to show frequency spaces not in use at the particular moment and therefore available for use, thus avoiding interference in crowded bands.

In military service, panoramic displays are widely used for quickly detecting hostile transmissions. Also in radio operations with other friendly forces panoramic displays permit one operator to guard a much wider frequency range and respond to incoming signals more quickly.

The panoramic principle has generally been applied in the form of a separate accessory or adapter to be used with a conventional receiver. One such adapter, called a "panadapter,"² is suitable for use with any receiver having an i.f. between 450 and 470 kc.

Panoramic adapters used by the services during the Second World War included the Navy model RDP for use with receivers having an i.f. of 30 Mc.³ An input of 25 μ v to the adapter at 30 Mc produces a $\frac{1}{4}$ -in. deflection on the scope after the gain control has been set for standard noise level. The maximum sweep of this model is ± 5 Mc. The sharp i-f amplifier in the adapter operates at 7.5 Mc. Other Navy models⁴ of the war have the following characteristics:

Designation	Input frequency,* kc	Max frequency sweep,† kc	Sharp i.f. of adapter, kc
RBV-1.....	400	50	226
RBV-1.....	400	100	226
RCX and RCX-1.....	450-470	200	226
RBW-2 and RBW-2M.....	5,250	1,000	912

* This is the i.f. of the receiver with which the adapter is used.

† The value shown is the total sweep to both sides of the center frequency.

¹ Moulie, W. E., Panoramic Principles, *Electronic Inds.*, July, 1944, pp. 86-88, 206. Williams, E. M., Radio-frequency Spectrum Analyzers, *Proc. IRE*, January, 1946, pp. 18P-22P.

² Popkin-Clurman, J. R., and B. Schlessel, Panoramic Reception, 1946, *QST*, March, 1946, pp. 22-27, 116, 118.

³ Publication PB-39310, Office of Technical Services, Dept. of Commerce, Washington. This is the instruction book on the Navy Model RDP. Microfilm copies and photostats are available.

⁴ Publication PB-39304, Office of Technical Services, Dept. of Commerce, Washington. This is the instruction book on the six Navy models shown in the tabulation. Microfilm copies and photostats are available.

The next step beyond using a panoramic presentation is to employ a detector to sound an alarm whenever any detail of the pattern changes. Such a detector¹ is necessarily elaborate but was produced as a laboratory development. This design has 40 control knobs for adjustment according to the pattern in view at any given time. After these controls are set, any change in the pattern will set off the alarm.

DIVERSITY RECEIVERS

Experience has shown that when a signal fades at an antenna, it is usually fairly strong at a distance of 10 or 20 wavelengths away. Diversity receiving systems² take advantage of this fact and consist of two or more spaced antennas with special receiving equipment arranged to deliver a single output signal of improved character.

The necessary spacing for the effective use of diversity reception, based on observations in England, is as follows:³

Frequency, Mc	Wavelength, m	Spacing of antennas, center to center, wavelengths	
		Best value	Min useful value
18.9	15.89	20	7
13.7	22	10	4
7.4	40.5	7-10	4

193. Diversity Receiver Details. The conventional arrangement of receivers operating in a diversity system provides for each antenna a complete receiver up to the point where the diode detectors are connected to a common load resistor. This resistor has an audio drop across it which for radiotelephone reception is the desired signal. The common load resistor also has a d-c drop which is applied as a v-c bias to all the receivers. With this arrangement, the d-c potential across the load resistor is determined by the strongest signal present. This is because this potential acts as a back bias on all the diodes and cuts off all of them except the one with the strongest signal. In this way only the strongest signal delivers output, and the noise that would otherwise come in from the other antennas is suppressed.

Practical points regarding receiver design for diversity operation evolved during years of experience may be found in the literature.⁴

On important circuits *frequency diversity* may be used in addition to the space diversity which has been considered so far. Frequency diversity merely consists of the transmission of the message simultaneously on more than one carrier frequency. An elaborate space and frequency diversity system between the United States and Alaska⁵ included four frequencies with dual space diversity for each one. This total of eight antennas was supplemented by one additional antenna for exploratory purposes. This system accommodated frequencies from 100 kc to 20 Mc, and was used for telegraph transmission.

Somewhat related to diversity methods is the "musa" antenna system developed about 1937,⁶ the term being coined from "Multiple-Unit Steerable Antenna." This

¹ Anderson, W. A., Alarm System for Panoramic Receivers, *Electronics*, July, 1946, pp. 92-95.
² Jelonek, Z., E. Fitch, and J. H. H. Chalk, Diversity Reception, *Wireless Engr.*, February, 1947, pp. 54-62.
³ Ladner, A. W., and C. R. Stoner, "Short Wave Wireless Communication," 4th ed., Chapman & Grimes, Inc., 1942; see pp. 464-468.
⁴ Lyons, W., Criteria for Diversity Receiver Design, *RCA Rev.*, June, 1947, pp. 373-378.
⁵ Ivers, J., CAA Alaskan Diversity Receiving System, *Communications*, January, 1946, pp. 40-43, 46; February, 1946, pp. 56-58, 61.
⁶ Friis, H. T., and C. B. Feldman, A Multiple-unit Steerable Antenna for Short-wave Reception, *Proc. IRE*, July, 1937, pp. 841-917; also in *Bell System Tech. J.*, July, 1937, pp. 337-422. Polkinghorn, F. A., A Single Sideband Musa Receiving System for Commercial Operation on Transatlantic Radio Telephone Circuits, *Proc. IRE*, April, 1940, pp. 157-170; also in *Bell System Tech. J.*, April, 1940, pp. 306-340.

consists of a group of rhombic antennas with phasing circuits so as to achieve (1) a high degree of vertical directivity and (2) the ability to control the vertical angle, thus "steering" the antenna. With this antenna system it is possible to favor the path on which the desired signal is arriving and discriminate against other paths. This is therefore a means of combating multipath propagation and its attendant selective fading. The *musa* has been found to give an improvement in quality for a considerable portion of the time and to afford a substantial improvement in signal-to-noise ratio under various conditions of poor transmission. However, the cost of the *musa* installation is substantial.

Diversity operation was found useful in the s-h-f range in tests at 4,500 Mc between San Diego and San Francisco.¹ At the receiving end of one of the relay links two parabolic antennas were used, one mounted 25 ft below the other. It was found that the short fades that occurred were not simultaneous at the two antennas. The outputs of the receivers were paralleled after the first video stage. The a-v-c voltages of the two receivers were connected together in the manner already described in this section. This system operated with pulse-time-modulated equipments.

In an f-m multiplex telegraph relay at 4,000 Mc, RCA found the use of two-antenna diversity advisable for hops exceeding 15 miles.²

WIDE-BAND RECEIVERS

In services such as television and pulse-type systems, receivers must possess a very much greater band width than required for simple radiotelegraph or radiotelephone operation. For example, band widths up to 4 Mc are necessary in television transmitters, and values up to 10 Mc or more characterize certain pulse transmissions.

194. **Gain Band-width Product and the Band Merit of Tubes.** Studies by H. A. Wheeler³ led to the realization that for the ideal coupling network, there is a maximum value of the product of stage gain and band width. This ideal maximum is

$$Af_w = \frac{g_m}{\pi \sqrt{C_v C_p}}$$

where A = voltage gain between equal impedances

f_w = band width, cps

g_m = transconductance of tube, mhos

C_v = total grid-to-cathode capacitance in circuit, farads

C_p = total plate-to-cathode capacitance, farads

If it is desired to realize a large part of this ideal maximum value, coupling networks of three or four reactances are necessary.

For a simple parallel-resonant circuit as the coupling means, the relation takes the form⁴

$$Af_w = \frac{g_m}{2\pi C}$$

where C is the total of C_p and C_v .

This expression for the maximum gain band-width product with a simple tuned circuit as the coupling has been used to obtain a figure of merit for comparing various tubes of interest as wide-band amplifiers. The point of view is that with wider and wider band widths, the obtainable gain with the particular tube declines until it reaches unity; *i.e.*, the output is no greater than the input. This makes A unity in the equation, so that it is only necessary to insert the values of g_m , C_v , and C_p for the particular tube to compute its band merit. Distributed capacitances, such as those of

¹ Huber, G. H., Space Diversity Reception at Super-high Frequencies, *Bell Lab. Record*, September, 1947, pp. 337-341.

² Gerlach, G. G., A Microwave Radio Communication System, *RCA Rev.*, December, 1946, pp. 576-600.

³ Wheeler, H. A., Wide-band Amplifiers for Television, *Proc. IRE*, July, 1939, pp. 429-438.

⁴ Ford, G. T., Characteristics of Vacuum Tubes for Radar Intermediate Frequency Amplifiers, *Bell System Tech. J.*, July, 1946, pp. 385-407; see pp. 388-390.

the coil and wiring, are omitted in this calculation, as they are not properties of the tube. The following table gives values of the band merit for several tubes:

Tube type	Transconductance, μmhos	Total of input and output capacitances, $\mu\mu\text{f}$	Band merit, Mc
6K7	1,600	20	13
6AG7	11,000	20	88
6AC7	9,000	16	90
6AG5	5,100	8.3	98
6AK5	5,100	6.8	120

An essential part of a wide-band receiver is the i-f amplifier. A general treatment of this is given Sec. 77. Stagger tuning¹ and feedback methods are used to obtain wide-band response.

After detection of the signal, it is usually necessary to amplify it further. In wide-band receivers such an amplifier is usually called a "video amplifier" because the signal is in many cases displayed on a scope as a television picture or as a radar pattern. For application to television see Chap. 19 of this volume and references;² for radar applications see references.³

RADAR

195. Wartime Receivers. Specifications of a number of radars used in the Second World War are available in the literature.⁴ Radars made for installation on the ground included models with carrier frequencies from 195 to 10,000 Mc, pulse durations from $\frac{1}{4}$ to 30 μsec , pulse repetition rates from 200 to 4098 pulses per sec, receiver noise figures from 7 to 18.5 db, and band widths from 1.25 to 10 Mc.

Air-borne radars included models operating at carrier frequencies from 2,500 to 10,000 Mc with pulse durations from 0.4 to 1.13 μsec , pulse repetition rates from 400 to 2,000 pulses per sec, receiver noise figures from 11 to 21 db, and receiver band widths from 1.9 to 8 Mc.

196. Maritime Radar Receivers.⁵ The FCC has assigned the following frequency bands for general maritime radar use: 3,000 to 3,246, 5,460 to 5,650, and 9,320 to 9,500 Mc. These bands accommodate the war designs in the neighborhood of 3,000 and 9,000 Mc and, in addition, establish the new 5,460 to 5,650 band at which experience can be gained. Radars for higher frequencies have better antenna directivity but

¹ Wallman, Henry, Stagger-tuned Amplifier Design, *Electronics*, May, 1948, pp. 100-105.

² Fink, D. G., Principles of Television Engineering, McGraw-Hill, 1940; see pp. 207-260. Video Amplifier H.F. Response, *Wireless World*, September, 1946, pp. 301-302; October, 1946, pp. 333-334. Schlesinger, Kurt, Low-frequency Compensation for Amplifiers, *Electronics*, February, 1948, pp. 103-105. Larsen, M. J., Low-frequency Compensation of Video-frequency Amplifiers, *Proc. IRE*, October, 1945, pp. 666-670. Wheeler, H. A., The Interpretation of Amplitude and Phase Distortion in Terms of Paired Echoes, *Proc. IRE*, June, 1939, pp. 359-385. Kallmann, H. E., R. E. Spencer, and C. P. Singer, Transient Response, *Proc. IRE*, March, 1945, pp. 169-195; includes many computed curves of response to unit-step signal. Sziklai, G. C., and A. C. Schroeder, Cathode-coupled Wide-band Amplifiers, *Proc. IRE*, October, 1945, pp. 701-711.

³ Schneider, E. G., Radar, *Proc. IRE*, August, 1946, pp. 528-578; see 551-556 for treatment of receivers. Berkner, L. V., Naval Airborne Radar, *Proc. IRE*, September, 1946, pp. 671-706. Ridenour, L. N., "Radar System Engineering," McGraw-Hill, 1947. Fink, D. G., "Radar Engineering," McGraw-Hill, 1947. Staff of Radar School MIT, "Principles of Radar," 2d ed., McGraw-Hill, 1946. Morrison, L. W., Jr., The Radar Receiver, *Bell System Tech. J.*, October, 1947, pp. 693-817. Miller, S. E., Considerations in the Design of Centimeter-wave Radar Receivers, *Proc. IRE*, April, 1947, pp. 340-351. *J. IEE (London)*, Part III A, No. 1, 1946, pp. 1115-1158; this issue gives four papers on radar receivers. Van Voorhis, S. N., "Microwave Receivers," McGraw-Hill, 1948. Other McGraw-Hill volumes related to receivers are Hamilton, Knipp, and Kuper, "Klystrons and Microwave Triodes"; Torrey and Whitmer, "Crystal Rectifiers"; Pound, "Microwave Mixers"; Chance, et al., "Electronic Time Measurements"; Solter, Starr, and Valley, "Cathode Ray Tube Displays"; Lawson and Uhlenbeck, "Threshold Signals."

⁴ Radar Specifications, *Electronics*, November, 1945, pp. 116-119. Airborne Radar Specifications, *Electronics*, February, 1947, p. 132.

⁵ Lynn, L. H., and O. H. Winn, Marine Radar for Peacetime Use, *Elec. Eng.*, May, 1946, pp. 271-273.

suffer greater attenuation due to rain. The choice of the best operating frequency is therefore a compromise.

Standards for marine radar¹ have been adopted by IMMRAN (International Meeting on Marine Radio Aids to Navigation) and by the Lakes Carrier Association, Cleveland, Ohio. The requirements of these specifications are generally satisfied by available commercial designs. In addition to these specifications, a monitor is required in order to serve as a check on the operation of the radar in service. The following are typical specification provisions:

Radio frequency*	3,000 or 9,300 Mc (approx)
Pulse duration, μ sec.	0.2-1
Pulses per sec.	750-3,000
Receiver noise figure, db max.	15
Receiver band width.	Not less than reciprocal of pulse duration
Minimum range of radar, yd.	100
Range resolution on shortest range scale, yd.	100
Azimuthal resolution.	1.6-2 deg for 9,300 Mc; 3.5-5 deg for 3,000 Mc
Antenna rotation, rpm.	6-15
Presentation.	PPI scope of 7 or 12 in. diam
Maximum range of radar, statute miles.	30-50
Number of range scales.	5 or 6

* Only 9,300 Mc is acceptable to Lakes Carrier Association.

SPECIAL RECEIVERS

Numerous cases arise in which designs of radio receivers for special purposes are needed. A few such instances illustrating the wide range of requirements are given.

197. Rural Radio Telephone.² The use of radio for rural telephone service has started on an experimental basis, employing frequencies in the region from 44 to 50 Mc with f.m. for the voice and a.m. for the ringing signals. Full deviation of the f.m. is ± 15 kc. Vertical polarization of the waves is used. The maximum distance covered in the work to date is about 20 miles.

The receivers in this service are crystal-controlled. At the telephone exchange a dual receiver is used, which brings in signals on both the regular frequency and on a special frequency which is provided for "reverting" calls (where one radio subscriber calls another in the same radio system). At the subscriber's station a transmitter which can operate on either one of two frequencies is provided for the same purpose. A subscriber's receiver, however, always operates on the same frequency.

198. Wired Radio. Telephone wires and power lines are extensively operated with "carrier" telephone signals at various frequencies in the v-l-f and l-f bands, the highest being about 150 kc. The receivers in these services are either t-r-f or superhet types.

In power-line practice, remote-control and telemetering operation is also obtained by wired radio. In such remote-control systems, the receiver may often be advantageously provided with larger than usual tubes for the diode detector and the preceding amplifier. In this way, the receiver can operate a moderate-sized relay as the first step in controlling the high-voltage power switching.

Coaxial-cable multiplex telephone systems operate up to much higher frequencies, such as 1 Mc or more.

199. F-m Terrain-clearance Altimeters.³ An f-m radar operates by rapid shift of the transmitter frequency, so that the returning echoes have a different frequency from the value then being radiated.⁴ Measurement of the difference between the outgoing and incoming frequencies can thus be made to indicate the distance of the reflect-

¹ Marine Radar Report, *Electronics*, July, 1947, pp. 144, 146, 244. Scroggie, M. G., Postwar Marine Radar in Great Britain, *Communications*, December, 1946, pp. 9-11, 41.

² Moore, J. H., P. K. Seyler, and S. B. Wright, Experimental Rural Radiotelephony, *Elec. Eng.*, April, 1947, pp. 346-348. Nylund, H. W., Rural Party-line Service by Radio, *Bell Lab. Record*, February, 1948, pp. 49-54.

³ Espenschied, L., and R. C. Newhouse, A Terrain Clearance Indicator, *Bell System Tech. J.*, January, 1939, pp. 222-234. F-m Radar Altimeter, *Electronics*, April, 1946, pp. 130-134.

⁴ Wolff, Irving, and D. G. Luck, Principles of Frequency-modulated Radar, *RCA Rev.*, March, 1948, pp. 50-75.

ing object. This principle has been extensively applied to radars designed specially to serve as terrain-clearance altimeters.

A commercial (RCA) f-m terrain-clearance altimeter, model AVQ-6, is generally similar to the wartime model AN/APN-1. The frequency is 440 Mc and the f.m. is at 120 cycles. The altimeter has two altitude ranges, one with a maximum of 400 ft and the other of 4,000 ft. On the 400-ft range the f-m deviation is ± 20 megacycles, producing an audio note of 19 cycles per foot of altitude. On the 4,000-ft range the deviation is ± 2 Mc producing a note of 1.9 cycles per foot of altitude. The maximum audio note on either range is 7,600 cycles.

200. Receivers with Privacy Feature.¹ For public radiotelephone service it is usual to provide some form of speech scrambling so that eavesdropping with simple types of receivers is prevented. One type of privacy equipment, described in Sec. 177, utilizes division of the voice-frequency range into several portions which are then exchanged, with or without inversion of the individual divisions. A simpler method of scrambling is to invert the entire speech band as a single unit, the usual practice being to have 1,550 cycles as the center point. In this case, 1,000 cycles, for example, becomes 2,100, and 2,100 becomes 1,000; similarly, 500 cycles and 2,600 change places; likewise, 100 and 3,000 cycles.

Warbling of the carrier frequency is also available as a privacy method. The extent of the warbling must obviously be kept within bounds in order to avoid excessive increase of the frequency spectrum.

201. Loran Receivers.² The loran system³ permits accurate navigational fixes to be obtained by planes and vessels by locating the intersection of certain hyperbolic curves. This method is based on the geometric fact that a hyperbola is the locus of points having a constant difference in distance from two fixed points. The loran system with carriers in the neighborhood of 1,900 kc furnishes fixes for ocean navigation at distances up to 750 miles during the day and 1,500 miles during the night. Four radio channels are in use, located at 1,750, 1,850, 1,900, and 1,950 kc. L-f service, first around 180 kc and now around 100 kc, is in the experimental stage. Slightly different pulse-repetition rates are used to distinguish several transmissions on the same frequency. There are eight of these rates in the low (L) group near 25 cycles, and eight additional ones in the high (H) group near 33 cycles.

The superior propagation characteristics of the l-f band have led to the development of the l-f loran system operating in the neighborhood of 180 or 100 kc. This system has been approved by ICAO (International Civil Aviation Organization) as a means of long-distance airplane navigation. Developmental work by the Watson Laboratories of the U.S. Air Force has indicated that a few such systems would cover a large area of the earth. In efforts to obtain the best accuracy with the l-f loran system, it has been found desirable that the receiver have a good band width, such as 20 or 25 kc, to permit the pulse to rise rapidly enough to distinguish the direct ground-wave signal from the sky wave arriving subsequently. The severe atmospherics sometimes encountered at this frequency suggest that variable selectivity might be desirable to permit narrowing of the band when needed.

UNDERWRITERS' REQUIREMENTS

Numerous provisions for the reduction of shock and fire hazards have been established by Underwriters' Laboratories, which is sponsored by the National Board of Fire Underwriters. Receiver models examined and found to comply with these provisions are listed under a "Reexamination Service," and the manufactured sets of these models are then privileged to bear a label showing approval by Underwriters' Lab-

¹ Reich, H. J., "Theory and Applications of Electron Tubes," 2d ed., McGraw-Hill, 1944; see pp. 299-300.

² The term "loran" is a coined word for long-range navigation.

³ Pierce, J. A., An Introduction to Loran, *Proc. IRE*, May, 1946, pp. 216-234; 2-Mc Sky-wave Transmission as Applied to Loran, *Electronics*, May, 1946, pp. 146-153. The Loran System, *Electronics*, November, 1945, pp. 94-99; December, 1945, pp. 110-115.

oratories. The latest edition of the "Standard for Power-operated Radio Receiving Appliances" of Underwriters' Laboratories will be furnished to receiver design engineers on request to any one of their offices in New York, Chicago, or San Francisco. This standard is a pamphlet giving the specific electrical and mechanical requirements for the approval of equipment, and covers radio receivers, television receivers, phonographs for home use, and recorders.

202. Receiver Hazards. With regard to fire hazard, one of the most important requirements for a-c receivers is that the power transformer must be completely housed in noncombustible material without louvers or holes except as required for the entry of wires. The purpose of the housing is to prevent the escape of flame or molten metal in case the transformer is overloaded by the short circuit of some secondary winding.

With a-c-d-c sets, the chief fire hazard is the first electrolytic filter capacitor, which for a-c operation and a short circuit of the rectifier tube is subjected to a severe condition. The connection of the filter circuit to a tap on the rectifier-tube filament is advantageous in reducing this hazard.

Shock hazard includes exposed conducting parts such as antenna lead, ground lead, metal cabinet if used, phono tone arm if used, and shafts and setscrews of control knobs.

Various requirements must be satisfied by flexible cords with regard to (1) character of wire, (2) bushing or equivalent treatment of the hole, (3) strain relief to prevent a pull on the cord being transmitted to internal connections, and (4) means to prevent the cord from being pushed into the hole.

Temperature limits for various materials inside the receiver must not be exceeded. Each wire must be suitable for the voltage, temperature, and other conditions prevailing.

A-c-d-c receivers with the chassis directly connected to one side of the power mains are permitted under certain conditions including the provision of a substantial back cover and interlock. In the case of f-m receivers for a-c-d-c operation, this has been found the most practical arrangement because it avoids the regeneration difficulties encountered with an isolated chassis.

COLOR CODES FOR WIRING

The RMA Standard REC-108 has established various color codes not only for components but for chassis wiring.

203. Chassis Wiring Color Code.

Black.....	Grounds, grounded elements, and returns
Brown.....	Heaters or filaments, off ground
Red.....	Power-supply B plus
Orange.....	Screen grids
Yellow.....	Cathodes
Green.....	Control grids
Blue.....	Plates
Violet.....	Not used
Gray.....	A-c power lines
White.....	Above- or below-ground returns, a.v.c., etc.

It is also specified that, when antenna and ground connections are provided on the receiver, the antenna lead shall be blue and the ground lead black. In the case of leads for special antenna connections, combinations of blue and black are specified.

204. Transformer Color Code. Power transformer leads have been standardized as follows:

Use	Color
Two-wire primary.....	Both leads black
Tapped primary.....	Start of winding, black; tap 50/50 black and yellow; finish, 50/50 black and red; if tap is not at center, it is to be nearer the finish end
Rectifier filament.....	Both leads yellow; tap if used, 50/50 yellow and blue

Use	Color
Rectifier plate.....	Both leads red; tap 50/50 yellow and red
Amplifier filament winding No. 1.....	Both leads green; tap if used, 50/50 green and yellow
Amplifier filament winding No. 2.....	Both leads brown; tap if used, 50/50 brown and yellow
Amplifier filament winding No. 3.....	Both leads slate; tap if used, 50/50 slate and yellow

For i-f transformers, the following colors have been standardized:

Use	Color
Plate.....	Blue
B+.....	Red
Grid or diode.....	Green
Grid return or diode return.....	Black
Secondary tap if used.....	Green and black

The colors of audio-transformer leads have been standardized as follow:

Use	Color
Plate.....	Blue
B+.....	Red
Other plate for push-pull, start of winding.....	Blue or brown
Grid or high side of moving coil.....	Green
Return or low side of moving coil.....	Black
Other grid for push-pull, start of winding.....	Green or yellow

The leads of speaker fields have been standardized as follows:

Use	Color
Start of winding.....	Black and red
Tap if used.....	Slate and red
Finish.....	Yellow and red

ANTENNAS FOR BROADCAST RECEPTION

A general treatment of antennas is given in Chap. 14, and loop antennas are discussed in Sec. 31, *et seq.* of the present chapter. The present section gives data on some open-type antennas especially designed for broadcast reception.

205. Master Antennas. Technique is available for the design of master antenna systems suitable for supplying many outlets, such as in apartment houses, and providing service on the broadcast band, the h-f international and amateur bands, and the lower television channels.¹

Installation of antennas for furnishing 100 Mc f-m service, as well as h-f and m-f reception, have been made on a large scale in some apartment-house developments.² One antenna in these installations supplies one antenna coupler which delivers signal to one or two transmission lines consisting of RG/59U flexible coaxial cable. Each of these cables supplies as many as 18 individual outlets. In this installation the outlets are directly in parallel across the transmission line; *i.e.*, no receiver couplers are used. Each run of RG/59U is terminated at the end with a 100-ohm resistor. A polarized outlet is used in each apartment to prevent a ground being placed on the center conductor of the RG/59U. A power outlet for the 117-volt supply of the radio receiver is provided conveniently at the same location.

206. F-m Antennas. Use of the home electric wiring as an antenna for f-m signals on the 100-Mc band has been found practical. This form of antenna, although not so good as a well-installed outside dipole, has been found an acceptable substitute and permits f-m table-model receivers to operate without the necessity of installing an outside antenna. A minor difficulty is that the results are sometimes poor at one outlet while they are good at another outlet in the same room.

In the 100-Mc f-m band extensive use has been made of folded-dipole antennas constructed of 300-ohm balanced transmission line and mounted in the top of the cabinet.

¹ Aceves, J. G., Recent Improvements in Master Antenna Systems, *Proc. Radio Club of America*, July, 1940, pp. 1-9.

² Klingenschmitt, F. A., FM Antennas for 12,497 Families in Metropolitan Life Projects, *FM & Television*, May, 1947, pp. 21-23.

Such an antenna extending across the width of a 40-in. cabinet, and extending along the sides for 9 in. had an efficiency within 1 db of a good dipole used as reference standard, and had voltage standing-wave ratios of about 5 at 88 Mc, 1.5 at 98 Mc, and 3.5 at 108 Mc. A more compact design, extending across the 24 in. width of a cabinet, and continuing for the full length of the 12 in. sides and then for 8 in. around the corner, thus having a total length of 64 in. had an efficiency of -2.5 db with respect to the reference standard, and voltage standing-wave ratios of 6, 3.6, and 7, respectively, at the three frequencies already mentioned.

References

Receiver Practice and Design:

- Dawley, R. L., editor: "The Radio Handbook," 11th ed., 450 pp., Editors and Engineers, Santa Barbara, Calif., 1947.
- Electronics Training Staff of the Cruft Laboratory, Harvard University: "Electronic Circuit and Tubes," 948 pp., McGraw-Hill, 1947. Radio receivers are discussed in Chap. 23, pp. 741-782.
- Hensler, W. W.: New Trends in Receiver Design, *Radio News*, June, 1948, pp. 46-47, 170-172.
- "Mallory Technical Manual" (also called "MYE Technical Manual"), 407 pp., P. R. Mallory & Co., Indianapolis, 1942. Includes chapters on first detectors and oscillators, half-wave and doubler power-supply systems, vibrators and vibrator power supplies, phonographs, push-button tuning, frequency modulation, and correction of appliances to stop radio interference.
- "Radio Amateur's Handbook," 1950 ed., American Radio Relay League, West Hartford, Conn.
- "RCA Receiving Tube Manual RC-15," 256 pp., Tube Dept., RCA, Harrison, N.J. Contains about 90 pages on general characteristics and uses of tubes.
- Rust, N. M., O. E. Keall, J. F. Ramsay, and K. R. Sturley: Broadcast Receivers: A Review, *J. IEE (London)*, Part III, June, 1941, pp. 59-96. Discusses numerous phases of receiver design.
- Smith, F. L.: "The Radiotron Designer's Handbook," 3d ed., 325 pp., Amalgamated Wireless Valve Co., Sydney, Australia, 1940. Available from RCA Mfg. Co., Harrison, N.J.
- Strutt, M. J. O.: "Modern Ultra- and Extreme-shortwave Reception," 387 pp., Van Nostrand, New York, and Macmillan, London, 1947.
- Sturley, K. R.: "Radio Receiver Design," Part I, 435 pp., Wiley, 1943. Covers r-f amplification and detection. Part II, 480 pp., published by Chapman and Hall, London, 1945, covers audio amplification, power supply, automatic gain control, special tuning methods, receiver measurements, f-m reception, and television reception.
- Van Voorhis, S. N.: "Microwave Receivers," 618 pp., McGraw-Hill, 1948.

Terminology in British Radio Literature:

- Accumulator = storage battery
 Bush = bushing
 Decoupling = use of filter to prevent feed-back coupling
 Earth = ground
 E.H.T. = extra high tension = plate supply for cathode-ray tube
 Former = coil form
 Four-pole = four-terminal passive network
 Gramophone = phonograph
 H.T. = high tension = plate supply
 Long-wave band = 900-2,000 m = 333-150 kc
 L.T. = low tension = filament supply
 Mains = power wires in the house for lighting and appliances
 Medium-wave band = 200-550 m = 1,500-550 kc
 Picofarad = pf = micromicrofarad
 Radiogram = phonoradio combination
 Relay = furnishing of programs over wires to homes or apartments from a master radio receiver or by wire from the studio of a radio station
 Screening = shielding
 Shroud = strap for holding transformer laminations
 Speech coil = voice coil (of a loud-speaker)
 Tag = lug
 Torch = flashlight
 Valve = tube
 Valve holder = tube socket

CHAPTER 18

RADIO BROADCASTING

BY CARL G. DIETSCH¹

1. Principal Elements of a Broadcasting System. All the equipment of a broadcasting system extending from the microphone to the radiating antenna of the radio transmitting station will be considered as part of the system. A general circuit layout of typical facilities of the kind used in the larger broadcasting centers for supplying a network of stations with program service is represented by the simplified diagram, Fig. 1. Equipment of a single studio is represented; that of other studios of the usual group would be similar and would be at the point marked on the program bus. Inasmuch as many programs, such as the broadcasting of special events, originate at remote points, in most cases a great distance from the studio, the layout of the facilities for remote pickups, sometimes termed "nemo" programs, has been included to illustrate the use of telephone lines as well as point-to-point radiotelephone communication to complete the circuits necessary.

A list of the essential elements of the system is as follows:

1. Microphones
 - a. Studio
 - b. Remote pickups
2. Apparatus for controlling and conveying microphone output
 - a. Studio control booth
 - (1) Preliminary amplifier
 - (2) Microphone mixers
 - (3) Studio amplifier
 - (4) Volume control or faders
 - (5) Volume indicator
 - (6) Monitoring speaker
 - b. Remote pickups
 - (1) Preliminary amplifier
 - (2) Volume controls or faders
 - (3) Volume indicator
 - (4) Monitoring equipment
 - (5) Radiotelephone or wire-line facilities for intercommunication
3. Master control-room apparatus
 - a. Volume controls
 - b. Studio amplifiers
 - c. Relays and switching apparatus
 - d. Network channel amplifiers
 - e. Volume indicator
 - f. Monitoring facilities
4. Telephone-line facilities to local radio transmitting stations and to distant radio transmitters connected to networks
5. Radio transmitter
 - a. Line amplifier or limiting amplifier
 - b. Volume controls
 - c. Volume indicator
 - d. Radio transmitter
 - e. Monitoring equipment
 - (1) Monitoring rectifier and speaker

¹ Engineering Department, RCA Communications, Inc.

- (2) Modulation-percentage indicator
- (3) Carrier-frequency monitor
- f. Antenna

2. A-f Range. Perfect reproduction of a sound transmitted through an electro-acoustic system requires that the system pass all the audible frequencies of the sound

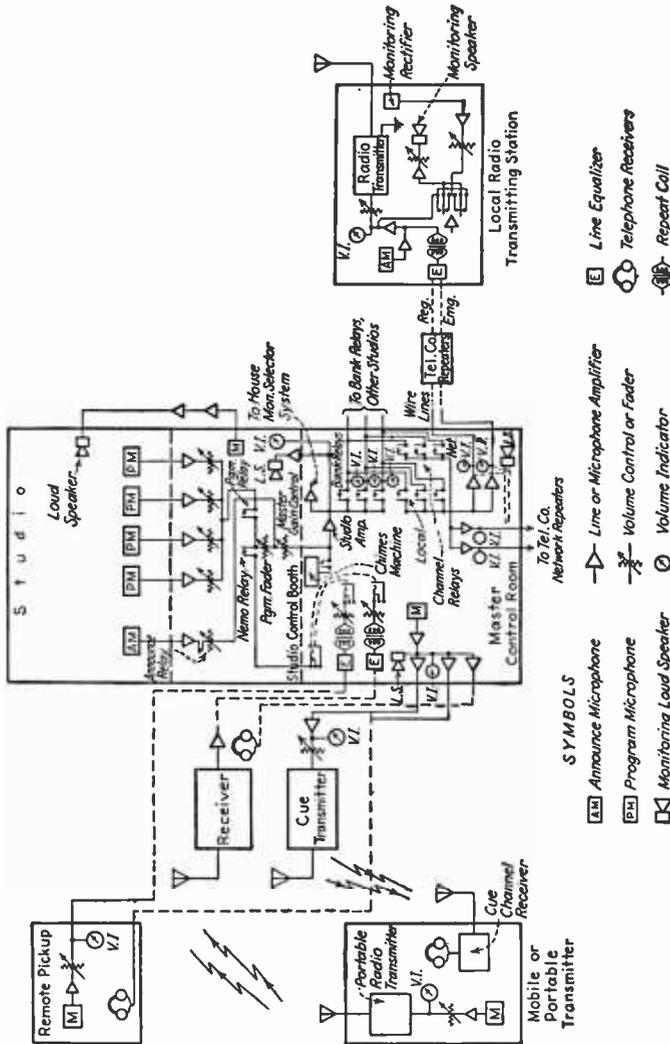
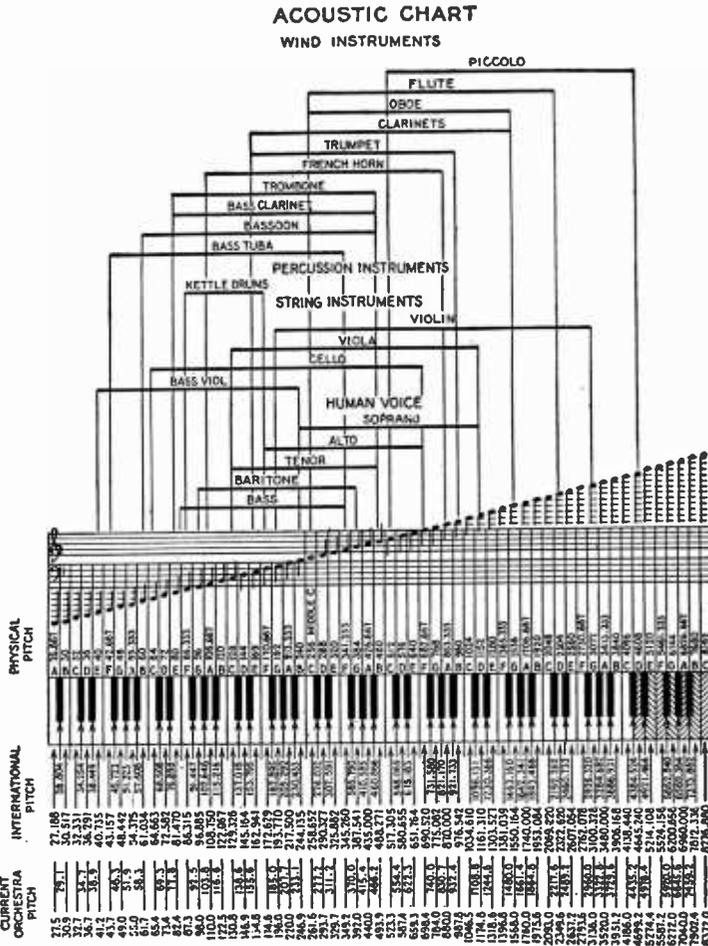


Fig. 1. Schematic diagram of typical broadcast transmitting system.

in their relative intensities. Under these conditions of reproduction, the listener would be conveyed acoustically from his loud-speaker to a point near the sound source except that the absolute sound level might be different from that experienced at the point of origin.

A correlated acoustic chart of the frequency range of various musical instruments

within the orchestral range and the different voices which constitute the vocal range is shown in Fig. 2. The shaded keys are not included on a standard piano keyboard. The extreme organ range not shown on the chart is from 16 to 16,384 cycles physical pitch. The extreme frequency-transmission ranges necessary to produce perfect



range, such as that which exists in broadcasting technique, an acceptable reproduction of the sound sources may be secured within a band width of between 30 to 9,000 cycles.

The engineering and economic limitations of the frequency range used for standard broadcasting lie in restrictions of the use of the upper audio frequencies due largely to a limited band width of the modulation spectrum contained between the presently assigned carrier frequencies of 10-kc separation,¹ for standard broadcasting in the spectrum between 550 and 1,600 kc.

Note	Cps	Organ pipe	Remarks
C ₁₀	33,488	Beyond limit of audibility for average person
C ₉	16,744	Considered ideal upper limit for perfect transmission of speech and music
	15,000	Considered as upper limit for high-quality transmission of speech and music
	10,000	Considered as satisfactory upper limit for high-quality transmission of speech and music
	9,000	Considered as satisfactory upper limit for high-quality transmission of speech and music
C ₈	8,372	¾ in.	Highest note on fifteenth stop
C ₇	4,186	Highest note of pianoforte
G ₆	3,136	
E ₅	2,673.2	Approximate resonant frequency of ear cavity
	3,000	Considered as satisfactory upper limit for transmission of speech for ordinary communication
C ₄	2,093	Maximum sensitivity of human ear
	2,000	Mean speech frequency from articulation standpoint
	1,500	
A ₄	880	Representative frequency of telephone currents
	800	
E ₄	659.3	
A ₃	440	Orchestral tuning (see note below)
C ₃	261.6	Considered as satisfactory lower limit for good-quality transmission of speech
	200	
C ₂	130.8	Considered as satisfactory lower limit of high-quality transmission of speech and music
	100	
E ₂	82.4	8 ft	Lowest note of cello
C ₁	65.4	
B ₀	61.7	
C ₀	32.7	16 ft	Lowest note of average church organ
	30	Considered ideal lower limit for perfect transmission of speech and music
A	27.5	Lowest note of pianoforte
G	24.5	
C	16.35	Lowest audible sound. Longest pipe of largest organ

Notes of the "gamut".....	C	D	E	F	G	A	B	C
Vibration frequencies proportional to.....	1	9/8	5/4	4/3	3/2	5/3	15/8	2
Intervals between successive notes.....	9/8	1 1/8	1 1/5	4/3	1 1/2	5/3	1 1/4	8

FIG. 5. Frequencies to be transmitted on a high-quality system. NOTE: Nearest note is indicated. Scale A = 440 cps based on middle C₁ (symphony pitch) = 261.6 cps.

An overlapping of the modulation frequencies of a "wanted" station by those of an "unwanted" station of 10,000-cycle separation restricts the range of frequencies to a broadcast listener usually considerably below that which is passed by the broadcasting system itself. The high quality of programs available from broadcasting facilities which have an over-all uniform frequency response from the microphone to the antenna within 2 db from 30 to 9,000 cycles and above, cannot therefore be appreciated by the average listener because of limitations in the average broadcasting-receiver frequency response and restrictions in the present standard broadcast band due to the

¹ Echersley, P. P., Minimum Frequency Separation, *Proc. IRE*, February, 1933, p. 195.

10-kc channel separation. Adjacent channel interference in the form of cross talk or "monkey chatter" prevents the satisfactory reception of the higher audio frequencies.

This limitation is not apparent in the case of f-m broadcasting, since the channels are assigned sufficiently far apart to permit transmission and reception of audio frequencies up to 15,000 cycles. This is obviously an advantage over standard broadcasting in that the entire audible frequency range may be satisfactorily transmitted and reproduced at the receiver by the f-m system.

Table 1. Peak Power of Musical Instruments

Instrument	Peak Power, Watts
Heavy orchestra.....	70
Large bass drum.....	25
Pipe organ.....	13
Snare drum.....	12
Cymbals.....	10
Trombone.....	6
Piano.....	0.4
Trumpet.....	0.3
Bass saxophone.....	0.3
Bass tuba.....	0.2
Bass viol.....	0.16
Piccolo.....	0.08
Flute.....	0.06
Clarinet.....	0.05
French horn.....	0.05
Triangle.....	0.05

3. **Volume Range.** Table 1 gives the peak power of various musical instruments playing triple forte. A violin playing very softly has an output of about 4 mw, whereas that of a full orchestra has a peak value of 70 watts. The intensity range of the sound sources in this case is about 43 db. Owing to limitations in broadcasting circuits, background noise, and the modulation capabilities of the transmitter, this volume range must be in most cases compressed within the limits which can be handled by the wire lines and their associated equipment, as well as the transmitter where serious amplitude distortion results if modulation peaks, except those of extremely short time duration, exceed the modulation capabilities of the transmitter.

STANDARD REFERENCE LEVELS

4. **Vu and Decibels.** The electrical signal intensity or level of an audio signal passing through each particular circuit of the broadcasting system, including the studio equipment, wire-line facilities, and the broadcasting station, must at all times be carefully adjusted. The adjustment must be such that the transmitter program signals will remain within the limits which every part of the system can transmit without objectionable distortion due to overloading or from interference caused by noise, such as the interference produced by cross talk, induction, rectifier ripple, etc., inherent in equipment as well as associated wire lines. A convenient and consistently accurate method of measuring the amplitude of the signals is required, as well as a reference level common to the entire system. From this reference level, termed the *zero reference point* or zero vu, is based the amplitude of the program waves throughout the system. The zero reference level of 1 mw is sometimes expressed as zero dbm where reference is given to the term db. It also serves as a reference level from which the amplitude of interference or noise may be measured in terms of a sine wave.

For broadcasting technique together with the interconnecting wire lines between studios and broadcasting stations there has been established a standard energy reference level of 1 mw. For the standard line impedance or pure resistance of 600 ohms at the terminals of a piece of apparatus in the system the zero reference level in vu would correspond to $\sqrt{0.6}$ rms volts of 1,000-cycle sine-wave electrical energy as measured by a standard a-c voltmeter across the terminals (see also Chap. 6).

Since program signals have wave shapes that are very complex and because peaks of

these complex waves are liable to cause overloading, a standard volume indicator¹ has been adopted for the purpose of measuring program levels at all parts of a broadcasting system so that the correct signal level can be maintained without objectionable overloading. This instrument (see Volume Indicators) is calibrated to read vu on a logarithmic scale. It has electrical characteristics approximately equivalent to those of an rms instrument. For signals having sinusoidal wave shape, the vu readings on this standard instrument should follow the decibel-voltage curve shown in Fig. 6. However, since the instrument is designed and used for measurement of complex program waves, the vu level of a particular program wave is as indicated by this standard volume indicator because of its particular characteristics. The term vu is therefore

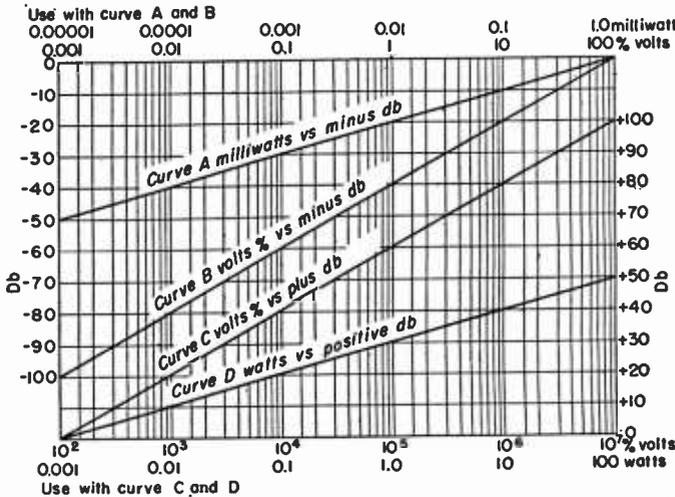


FIG. 6. Relation between decibels and watts or per cent volts. Zero level is 1 mw across 600 ohms.

associated with the reading of this meter whereas the term decibel follows steady-state conditions and mathematical laws.

Simultaneous with the establishment of the reference level of 1 mw the former 0-db level of 12.5 mw was abandoned and the value for standard apparatus and telephone-line termination impedances for broadcasting was changed from the previous value of 500 ohms to the present standard 600 ohms.

AUDIO FACILITIES

5. Microphone Requirements. By means of the microphone, acoustic energy of sound waves produced for broadcasting purposes is converted into electric energy, the wave shape of the latter conforming to that of the sound waves. The principal requirements of a microphone which will produce high-quality conversion are as follows: a relatively high sensitivity with respect to its inherent noise level, a uniform wave response over the frequency range desired, a substantially uniform frequency response over the angles included by its directivity characteristic, and mechanical and electrical ruggedness.

With some reservation, one may say that all forms of acoustoelectric transducers require the introduction of an obstacle into the path of the sound waves. To be effective, the active element of a microphone must either partake of, or otherwise influence

¹ Chinn, H. A., D. K. Gannett, and R. M. Morris, A New Standard Volume Indicator and Reference Level, *Proc. IRE*, January, 1940.

the motion of the air particles, or it must respond in some way to the pressure variations on its surface. Some portion of the instrument such as the outside case, regarded as a rigid obstacle, must reflect some of the incident wave energy, whereas the element which responds to vibration from the sound waves must reradiate some of the energy exciting it. An instrument of high sensitivity and efficiency must, therefore, absorb a considerable proportion of the sound energy reaching it and convert it into electric energy. Faithful reproduction, therefore, is dependent upon the physical size and shape of a microphone. These features enter into the distortion of the true sound field, as well as the characteristics of the elements used to convert the sound into electric energy with a minimum of wave distortion.

Inasmuch as the quality of reproduction of speech and music is dependent upon the acoustic properties of the room¹ containing the sound sources and the placement of the

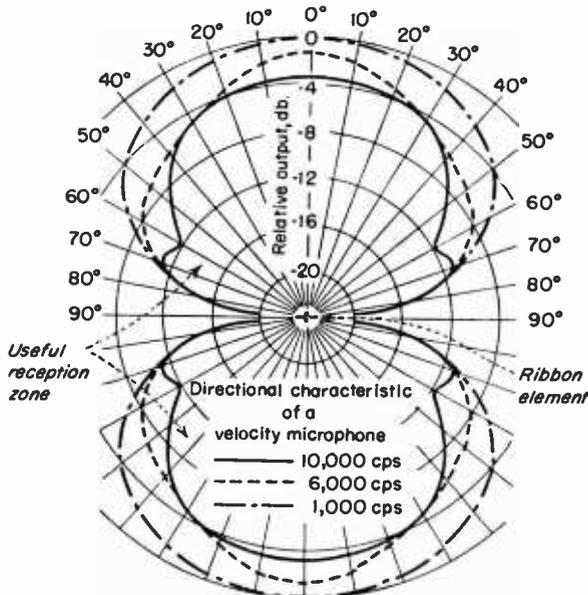


FIG. 7. Directional characteristics of RCA velocity microphone.

microphone with respect to them, satisfactory results while using even the best instruments require a knowledge of the technique of microphone placement.

6. Velocity Microphone. This instrument gets its name from the movement of a metallic ribbon under the motion of air particles impinging upon it, thus setting up by electromagnetic induction an emf corresponding to the amplitude variations of an incident sound wave.

The commercial form of the RCA type 44BX² consists of a thin metallic ribbon suspended between the poles of powerful permanent magnets with the ribbon length perpendicular to, and its width in the plane of, the magnetic lines of force. It is moved from its position of equilibrium by the difference of pressure between its two sides. This pressure difference between the front and back of the ribbon is the same as that produced in a sound field between two points in space separated by this distance. The pressure difference between the front and back of the ribbon is proportional to fre-

¹ Hanson, O. B., and R. M. Morris, Design and Construction of Broadcast Studios, *Proc. IRE*, **19**, January, 1931.

² Olson, H. F., *J. Soc. Motion Picture Engrs.*, **16**, 695, 1931; *J. Acous. Soc. Am.*, **3**, 56, 1931.

quency. Since the acoustic impedance of the system is also proportional to frequency and since velocity in a mechanical system is the ratio of the pressure to the acoustic impedance, the velocity of the ribbon is independent of frequency.

With a ribbon constructed to have a natural period below the audible range, the frequency response is free from severe irregularities prominent in some pressure-operated types because of cavity and diaphragm resonance and from pressure-doubling effects produced at the higher frequencies. The ribbon is made light enough so that its motion will conform with the motion of air particles even at very high frequencies, with a result that the response of the velocity microphone is uniform over a wide range of frequencies.

The velocity-type microphone is markedly directional. With a plane-progressive wave the response in front and back of the instrument varies with the cosine of the angle between the direction of the sound wave and the normal to the ribbon. Since these directional properties are practically independent of frequency, they become useful in discriminating against undesired sounds and for obtaining a desired relation between the sounds from different sources and from reverberant sound in a studio. Its response¹ to reverberant or reflected sound is one-third that of a nondirective system, with the result that it can be used at a distance from a sound source of 1.7 times the distance of a nondirective type and still give the same results with respect to undesired reverberant sounds.

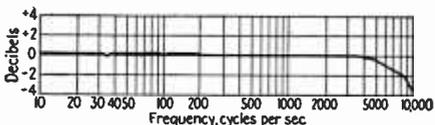


FIG. 8. Velocity-microphone characteristics.

Because of the directional properties of the instrument, its sensitivity is at a maximum in directions in front and back perpendicular to the plane of the ribbon. With an input sound pressure of 1 dyne per sq cm the unit will normally deliver open circuit across the 250-ohm tap an output level of -74 db compared to a zero level of 1 volt.

7. Moving-coil or Dynamic Microphone. This type of instrument (such as the Western Electric 630A) utilizes the light movable coil contained in a magnetic field to

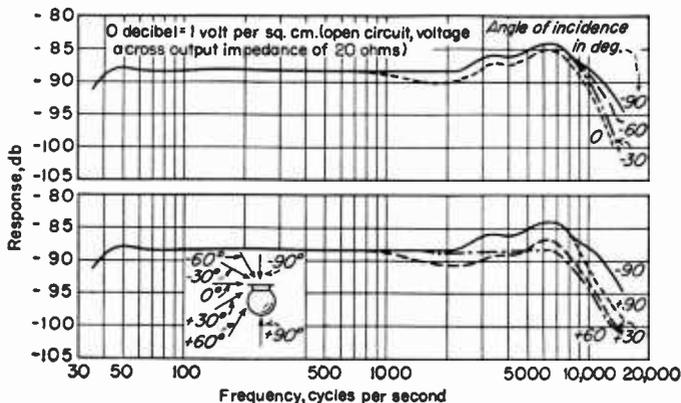


FIG. 9. Field calibration of dynamic microphone showing effect of angle of incidence of the sound wave.

produce an emf which conforms with the sound waves impinging upon the dome-shaped diaphragm.

The assembly is composed of a coil of fine aluminum ribbon edgewise wound and attached rigidly to a duralumin diaphragm of low mechanical stiffness which supports

¹ Olson, *loc. cit.*

the coil in a radial magnetic field of a permanent magnet made from high-grade magnet steel. The diaphragm has a rigid dome-shaped center and a tangentially corrugated annulus. It has a high area/stiffness ratio. The diaphragm is cemented to a raised annulus on the outer pole piece. The outer and inner pole pieces are of soft iron and are welded directly to the magnet. The diaphragm is damped by an acoustic resistance which is supported below the coil by a brass ring, which in turn is held in place by rubber gaskets.

When the diaphragm vibrates in response to the sound waves impinging upon its surface, the coil vibrates in a like manner and cuts the magnetic lines of force. The emf generated in the coil is substantially proportional to the sound vibrations which cause the diaphragm movement.

The spherical shape of the microphone housing and its size are such that the housing fits closely over the diaphragm and thus produces very little more diffractive effect than the diaphragm itself. To prevent resonance within the spherical case an acoustic resistance baffle is provided to divide the space in two parts. A tube with its outlet at the back of the housing serves the double purpose of equalizing the inside and atmospheric pressures and of increasing the response of the instrument at low frequencies.

This microphone was designed to provide a uniform frequency response in all directions and has been termed a *nondirectional microphone*.¹ For this reason the small spherical shape was selected as well as the method of mounting the diaphragm in a horizontal plane. A protective grid is provided over the diaphragm to control the resonance of the cavity in front of the diaphragm. This grid is most useful in the improvement of the frequency response of the instrument at frequencies from 8,000 to 15,000 cycles.

Wave-response calibration curves of this type of instrument indicate that the frequency characteristics are influenced to some extent at the higher frequencies by the angle of incidence from which the sound waves approach the diaphragm. Since the diaphragm is mounted horizontal, the instrument is entirely nondirectional with respect to the vertical axis.

In spite of the small physical size necessary to provide the nondirectional characteristics, the sensitivity is about -88 db where 0 db is equivalent to 1 volt per dyne per sq cm (open-circuit) voltage across the microphone output impedance of approximately 20 ohms.

The nondirectional characteristics of this microphone make it useful as a pickup for large orchestras and choruses where in most cases the sound arrives at the microphone from all directions. Unless the microphone response is uniform in all directions, distortion occurs due to discrimination against certain frequencies with directivity.

8. Unidirectional Ribbon Microphone. In certain forms of studio technique it is desirable to eliminate the pickup of unwanted sound in the rear of the microphone, such as audience noise, room echo, etc. Here the unidirectional microphone is very useful. The unidirectional instrument utilizes a lightweight corrugated ribbon suspended in a magnetic field in somewhat the same manner as the bidirectional velocity microphone, except that the ribbon is divided into two individual sections, one of which is pressure operated and the other velocity operated.

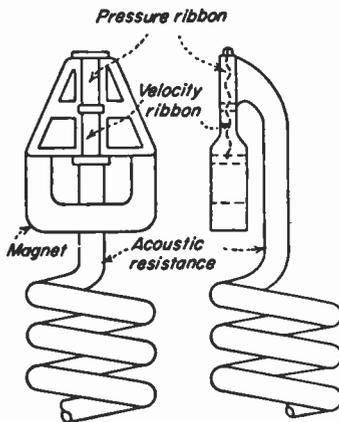


FIG. 10. Unidirectional ribbon microphone elements.

The field response of the pressure-operated section is very nearly uniform in all directions and may be expressed as $E = E_0$, whereas the response of the velocity section is bidirectional and is equivalent to $E = E_0 \cos \theta$. Since the sensitivity of the

¹ Marshall, R. N., and F. F. Romanow, *Bell System Tech. J.*, July, 1936, p. 405.

nondirectional pressure section is made to equal the greatest sensitivity of the bidirectional or velocity section, the combined polar field response characteristic of the two is equivalent to $E = E_0(1 + \cos \theta)$. In three-dimensional space, this is very nearly equivalent to a cardioid of revolution. The point of maximum sensitivity is directly to the front of the instrument, while directly to the rear the sensitivity approaches zero.

A uniform frequency response in all directions for the pressure-operated ribbon section is approached by allowing the operating face to be freely accessible to the atmosphere while the other side is terminated in an acoustic impedance very nearly equivalent to that of a very long pipe. Since a long pipe is too cumbersome for practical purposes, a short pipe of correct cross section, provided in coiled form, and loaded through its length with absorbing material, such as tufts of felt, exhibits a suitable acoustic resistance over a frequency range covering all but the lowest frequencies.

The operating properties of the velocity-actuated ribbon section are the same as were described previously for the bidirectional velocity microphone. The ribbon is driven from its equilibrium position by a difference of pressure between the two sides; the pressure difference being due to the difference in phase between the two sides. The vibration of the ribbon caused by the sound waves impinging upon it causes an induced emf to be generated in the ribbon. The directional characteristics of the ribbon section are practically independent of frequency.

The RCA 77-B unidirectional microphone has an open-circuit output level of approximately -81 db based on 1-volt zero reference level for a sound pressure of 1 dyne per sq cm at the 250-ohm output open-circuited impedance.

The Western Electric 639-AA cardioid directional microphone utilizes a ribbon element of special design in combination with a compact pressure type nondirectional element to secure a field response having a directional characteristic similar to a cardioid.

The pressure element is of the dynamic type having a dome-shaped diaphragm and constructed in some respects similar to the Western Electric 630A microphone previously described. Commercial instruments of this general type have switches which enable the directional characteristics to be changed at will.

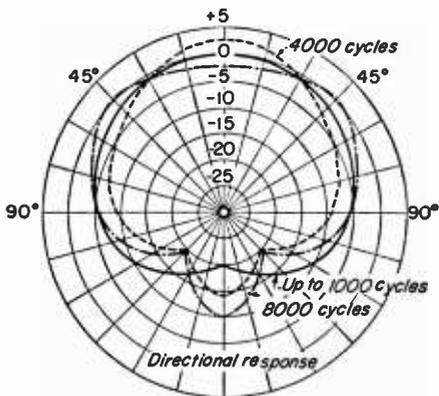


FIG. 11. Characteristics of RCA 77B unidirectional microphone.

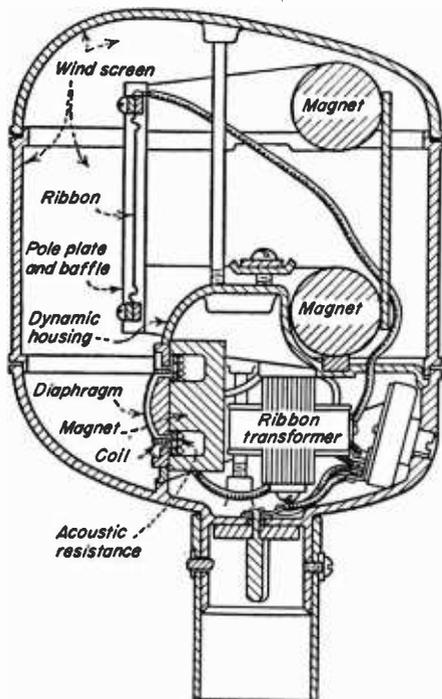


FIG. 12. Simplified cross-section view of the cardioid directional microphone.

9. Polydirectional Microphone. The RCA type 77D microphone combines the features of the velocity and pressure microphones to provide polydirectional characteristics by means of simple internal adjustments. This microphone is a high-fidelity instrument containing a simple ribbon suspended in an air gap formed by the pole pieces of a permanent magnet. The polydirectional microphone differs from the conventional velocity microphone in that one entire side of the microphone ribbon is covered by a connector tube which is coupled to a damped pipe or labyrinth. Behind the ribbon in the connector tube an aperture in the form of an adjustable shutter is

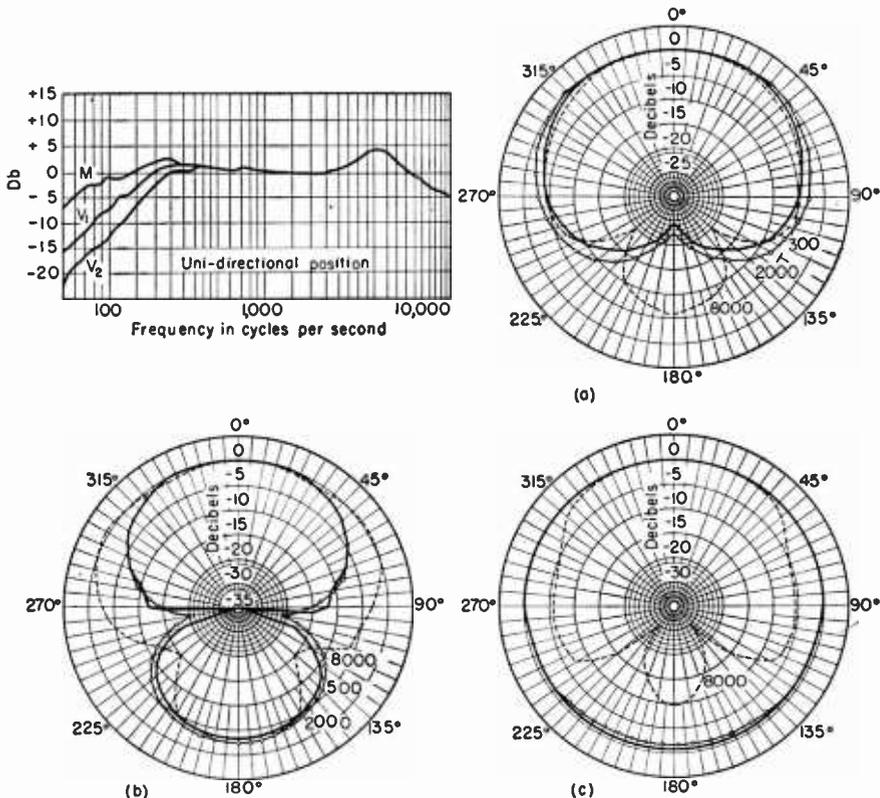


FIG. 13. Directional characteristics of the 77D microphone.

provided to permit various areas of opening between the ribbon and outside air. When the aperture is completely closed by the shutter, the microphone operates in the same manner as a pressure type. Under these conditions it has nondirectional characteristics. When the shutter is in the wide-open position, the ribbon is open to the air in front as well as to the rear and the microphone operates similar to the velocity type already mentioned. Under these conditions it has bidirectional characteristics. A common cardioid or unidirectional pattern is secured by an intermediate position of opening of the shutter. Directional patterns of the three types mentioned above are shown in Fig. 13. These are secured by settings of an adjustment screw on the rear of the microphone housing to marked positions.

To provide different amounts of l-f attenuation of the microphone response characteristics under conditions of different directional settings of the shutter, a reactor is

provided to shunt the output of the microphone transformer. A switch is provided within the lower shell to cut the reactor into the circuit as indicated by Fig. 14.

This microphone, by virtue of the polydirectional characteristics, provides a means of control in the ratio of directional to reverberant sound as well as to reduce unwanted

There are numerous applications of various directional patterns and the different response characteristics of this microphone to provide usefulness under various operating conditions. It is possible to secure any type of limacon¹ directional characteristic by means of the shutter adjustment. Because of the variable directional characteristics with respect to the sound source at a point directly in front of it, the sensitivity is dependent upon the directional setting. For bidirectional operation, the sensitivity is approximately equal to that of the velocity microphone previously described.

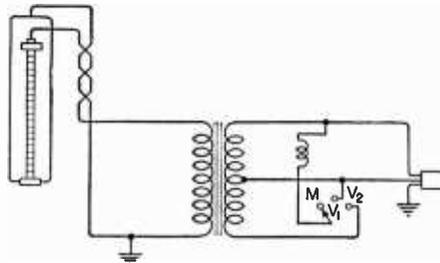


FIG. 14. Type 77D microphone connections.

10. Crystal Microphone. This microphone utilizes the piezoelectric phenomenon produced in plates cut from piezoactive crystals. Thin plates cut from Rochelle-salt crystals are used almost entirely for the elements of crystal microphones. In comparison to other crystalline piezoelectric materials, such as quartz, Rochelle salt exhibits greater sensitivity for this purpose and it responds quite readily to mechanical vibrations.

Crystal microphones may be classified under two individual groups: (1) those utilizing multiple sound cells in free space and (2) those utilizing bimorph crystal elements each excited mechanically by an associated diaphragm. In the first of these types utilizing the Brush Development Company assembly, termed the *sound cell*, the elements are plates having dimensions $\frac{3}{8}$ by $\frac{3}{8}$ by 0.30 in. cut from Rochelle-salt crystals along axes in such a manner that their inherent characteristics tend to cause elongation or contraction when they are subject to an electric field provided by foil electrodes.

By cementing together two such piezoactive plates which have tendencies to act in opposition to each other when a voltage is applied, an assembly is produced with a motion analogous to the mechanical motion of bending a bimetallic thermostatic strip acted upon by variation of temperature. The assembly consists of two plate combinations mentioned above, separated by an air space and held in position by a suitable mounting.

The cell is covered over with a membrane which serves as a pressure seal and to protect the crystals from the outside atmosphere. When the cell is placed in a sound field, pressure acting normal to the outer surfaces of the plates tends to cause bending, with a result that an emf is generated between the foil electrodes. The two plate combinations are connected in parallel. The wave form of this emf conforms with that of sound waves. Because of the small physical dimensions of the plates the frequency of mechanical resonance of the system is rather high, with the result that frequency response is quite uniform over a wide frequency range. Some models are quite uniformly sensitive up to 15,000 cycles.

Commercial models contain series and series-parallel groups of these sound cells ranging from 2 to as many as 24. The sensitivity of a single sound cell is approximately -90 db, while a multicell microphone has a sensitivity as great as -68 db.

The output impedance (which is purely capacitive) of these instruments is quite high. This sometimes requires them to be operated directly into the grid of an amplifier tube having a grid leak of about 5 mcgohms. The small physical dimensions of

¹ Olson, H. F., Polydirectional Microphone, *Proc. IRE*, February, 1944, p. 77.

² Marshall, R. N., A Cardioid Microphone, *Bell Lab. Record*, July 1939.

a single cell make it practically nondirective. This property is also characteristic of multicell units. Figure 15 shows an RC amplifier suitable for use with such a microphone.

The diaphragm type, such as the Brush Model AP utilizes a hermetically sealed bimorph crystal supported at three points within the microphone housing. Projecting to the center of the specially treated fiber diaphragm is a small drive pin. This engages the remaining corner of the bimorph crystal. Inasmuch as the bimorph crystal is highly sensitive in converting fluctuating mechanical stresses, such as those

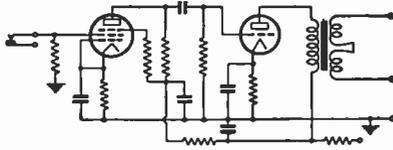


FIG. 15. Amplifier for use with crystal microphone.

caused by bending, into corresponding electrical fluctuations, the fluctuations in pressure created by the sound waves impinging upon the microphone diaphragm result in corresponding voltage fluctuations to be produced across the terminal ends of the bimorph crystal plates.

Similar to most pressure-operated microphones, this instrument is semi-directive. However, the smallness of the instrument assists in securing a rather uniform frequency response with direction. By placing the diaphragm facing the ceiling of the room, the instrument is essentially nondirectional in a plane through the diaphragm and parallel with the ceiling.

The output impedance of the crystal element is approximately 250,000 ohms. This permits long cables to be employed when either high or low impedance connections are used. By means of a high-quality matching transformer this microphone can be satisfactorily operated into circuits exhibiting 50, 200, or 600 ohms impedance. The frequency response of this instrument is substantially flat from 100 to 5,000 cycles. It has a variable control to allow manual adjustment of frequency-response characteristic while in operation. It has an output level of approximately -48 db based on a zero reference level of 1 volt per dyne per sq cm.

The crystal microphone shunted across the input resistor of the amplifier tube attenuates the higher frequencies of background noise such as those caused by thermal agitation. The properties of the crystals themselves are such that they are liable to damage, such as a change of frequency characteristics and output if the instrument is subjected to temperatures in excess of 120 to 125°F particularly for periods of several hours.

11. Condenser Microphones. The condenser microphone utilizes the principle of mechanical variation of thickness of the air dielectric of a charged electrostatic capacity as a medium to change acoustic energy into electrical energy of corresponding wave shapes. One form of this microphone consists essentially of an electric capacitor formed by a thin, tightly stretched duralumin diaphragm spaced approximately 0.002 in. and insulated from a flat brass disk called the *back plate*.

A polarizing potential difference is applied between the electrodes formed by the diaphragm and the back plate. The varying pressure upon the very thin diaphragm by the sound waves causes the electrostatic capacitance to vary by an amount in the order of 0.01 per cent of its normal value of 200 $\mu\mu\text{f}$.

The microphone has an aluminum alloy diaphragm 0.001 in. in thickness. The edges are clamped between threaded rings, the requisite stiffness being obtained by advancing the stretching ring until the desired resonant frequency, usually about 5,000 cps, is obtained. The space between the diaphragm and the back plate is hermetically sealed to prevent dust and moisture from entering and resulting in noise. The thin rubber auxiliary diaphragm, together with a small air-vent hole in the center of the back plate, is provided as an equalizing system for changes in atmospheric pressure.

On account of its inherent high-impedance characteristics, it is usual to incorporate an amplifier in the microphone housing to reduce to a minimum the length of the lead and the corresponding shunting capacitance between microphone and associated amplifier grid. Sometimes a compact amplifier is placed on the floor alongside the

microphone, the two being connected with low-capacitance cable. A d-c polarizing voltage in excess of 180 volts has been used, but this should never exceed 500 volts.

Developments upon the early Wente¹ models by using duralumin as a substitute for steel as diaphragm material brought the sensitivity of modern instruments to about ten times that of early models.

Since this is a pressure-operated type, there are inherent irregularities in its characteristics from acoustic and mechanical phenomena. The microphone diaphragm is subject to certain resonance frequencies as well as the cavity. These tend to disturb the smoothness of the response characteristic. The sound waves striking and being reflected from the flat surface of the diaphragm cause pressure doubling especially at high frequencies.

Below 500 cycles this instrument is practically nondirective whereas at frequencies above 2,000 cycles the directivity is very noticeable. This directivity has a tendency to discriminate against h-f noise and reverberation, and, under certain conditions where the studio does not accentuate the low frequencies, it has an advantage since the human ear responds more easily to background noise of higher frequencies than to lower frequencies.

The sensitivity of the microphone on the basis of an input sound pressure of 1 dyne per sq cm is approximately -60 db below 1 volt as measured at the output of the preamplifier.

The Western Electric 640A miniature condenser microphone unit² contains a diaphragm a fraction of an inch in diameter. The capacitance unit is mounted in one end of a tapered shell housing, of dimensions approximately 2½ in. in diameter and 7 in. long, which also contains the preamplifier. The weight of this microphone and preamplifier unit is 1¼ lb.

The output level of the complete instrument is -61 db below zero level of 1 volt per dyne per sq cm (open circuit) at the preamplifier output impedance of 50 ohms. Published field-response curves indicate a rather uniform frequency-response characteristic from 40 to 10,000 cycles with some dropping off of the higher frequencies as the angle is increased from an axis line normal to the diaphragm.

12. Carbon Microphones.

These devices use the variation resistance of carbon granules to produce electric waves from sound waves. A typical example of a "double-button" carbon microphone is shown in Fig. 16. The diaphragm of the microphone is made from duralumin 0.0017 in. in thickness and is clamped securely around its outer edge. Stretching of the diaphragm to give the desired resonant frequency, usually about 5,700 cycles, is done in two steps by means of two stretching rings. To ensure uniformly low contact resistance, the portions of the diaphragm which are in contact with the granular carbon are covered with a thin film of gold deposited by cathode sputtering. The carbon granules will pass through a screen having 60 meshes per inch but will be retained on a screen having 80 meshes per inch. Each button contains about 0.06 cc of carbon corresponding to about 3,000 granules.

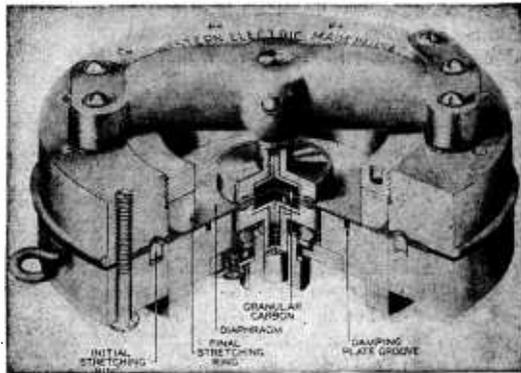


FIG. 16. Carbon microphone.

¹ Wente, E. C., *Phys. Rev.*, 19, 498, 1922.

² Harrison, H. C., and P. B. Flanders, An Efficient Miniature Condenser Microphone System, *Bell System Tech. J.*, July, 1932, p. 451. Hopper, F. L., *J. Soc. Motion Picture Engrs.*, September, 1939, p. 278.

The use of an air-damped stretched duralumin diaphragm has resulted in uniform response over a wide range of frequencies.

The operation of a carbon microphone may be affected by cohering (sometimes called *caking*) of the granules. Severe cohering causes a large reduction in resistance

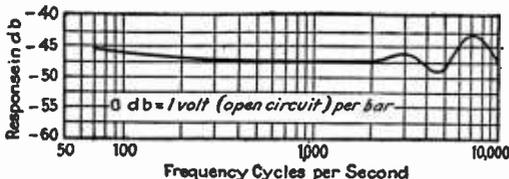


FIG. 17. Response of air-damped duralumin diaphragm.

and sensitivity which persists for an extended period unless the instrument is tapped so as to agitate mechanically the granules. One of the common causes of cohering is breaking the circuit when current is flowing through the microphone. Experience has shown that the use of a simple filter consisting of two $0.02 \mu\text{f}$ capacitors and three coupled coils, each having a self-inductance of 0.0014 henry, will effectively protect the microphone button without introducing an appreciable transmission loss; a potentiometer switch also serves to prevent caking.

The quality of transmission obtained with a double-button carbon microphone compares favorably with that secured with a condenser microphone; the carbon microphone has the disadvantage, however, of a high noise level or "microphone hiss."

Figure 18 shows the manner in which the carbon microphone is connected to its associated amplifier. The current through each button is usually in the neighborhood of 10 to 20 ma.

The sensitivity of the carbon microphone is somewhat higher than the other types. The average sensitivity is about -40 db.

Wave-response curves¹ for a carbon microphone show that response at normal incidence is quite uniform from 60 to 1,000 cycles. Above 1,000 cycles it increases rapidly, becoming about 15 db higher at 2,500 than at 1,000 cycles. This increase extends rather uniformly from 2,500 to 6,000 cycles, where there is a marked falling off.

13. Parabolic Reflector Microphone. The use of a large concave reflecting surface mounted behind a microphone has been found to give the instrument pronounced directional characteristics in the reception of sound waves. The system gets its name from the shape of the reflecting surface, a cross section of which contains a section of a parabola. By virtue of the microphone placement at the focus of the parabola of revolution or hollow paraboloid section, the sound waves striking the reflecting surface are concentrated upon that microphone diaphragm facing the inside of the paraboloid resulting in increased sensitivity of the instrument in line with the axis inside of the paraboloid.

The use of the reflector, therefore, makes possible the placement of the instrument sufficiently far from the sound source so that it is practically equidistant from all the instruments or voices, with a result that the problem of securing proper balance and volume control is simplified. The directional characteristic makes it possible to swing the microphone and its reflector as one would a searchlight and in this manner follow the action on the stage of an auditorium or on the field of a sporting event. There is an increase in sensitivity along the line of axis of about 4 to 1, owing to the use of the parabolic reflector.

Since the reflector increases the sensitivity and makes it possible to locate the microphone at a greater distance from the source of sound, it is desirable that the output of the microphone should fall off rapidly if the sound originates at a point displaced more

¹ Ballantine, Stuart, High-quality Broadcasting, *Proc. IRE*, 32, 576, May, 1934.

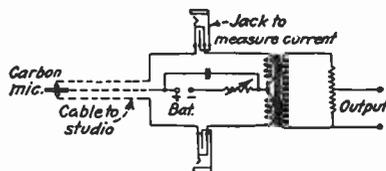


FIG. 18. Carbon microphone connections.

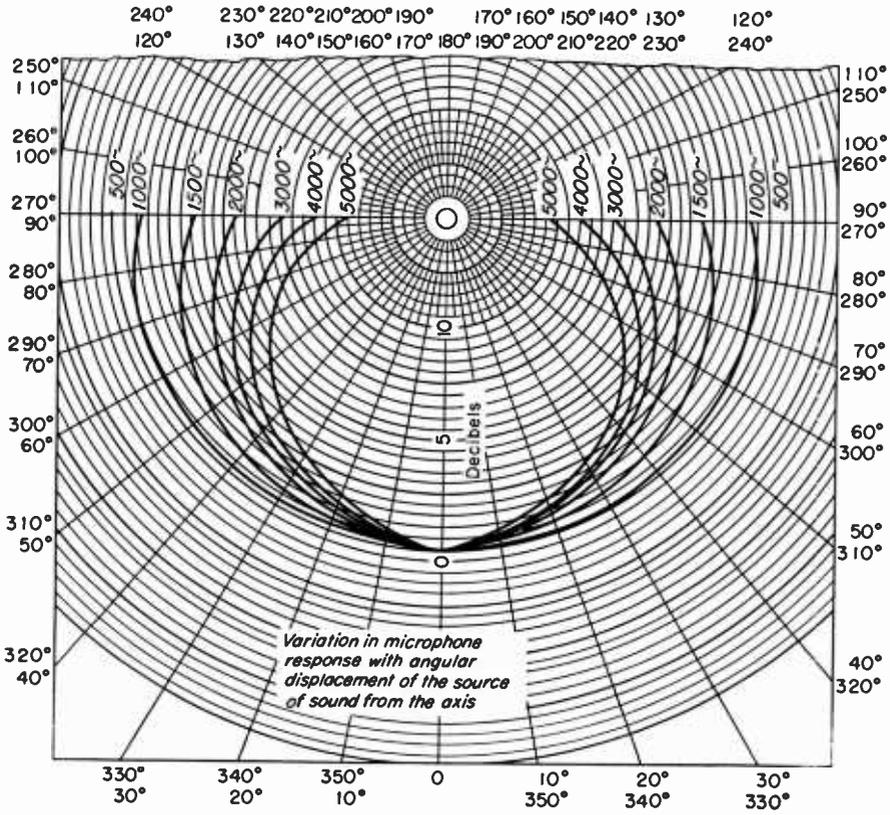


FIG. 19. Directional characteristic of carbon microphone.

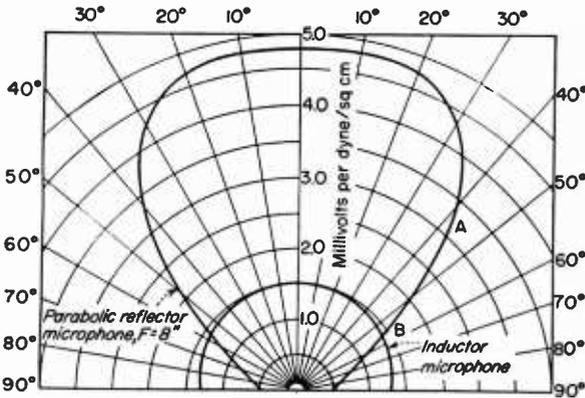


FIG. 20. Comparative axial response at 1,000 cps in millivolts per dyne per sq cm: (A) parabolic reflector; (B) inductor microphone.

than 30 deg from the axis of the instrument; if this characteristic is obtained, reverberation and reflections in the studio or auditorium will have very little effect.

The h-f response may be increased by as much as 15 db over the response at low frequencies by varying the position of the microphone in the reflector. However, in focusing the microphone, care must be taken to select the most useful frequency range,

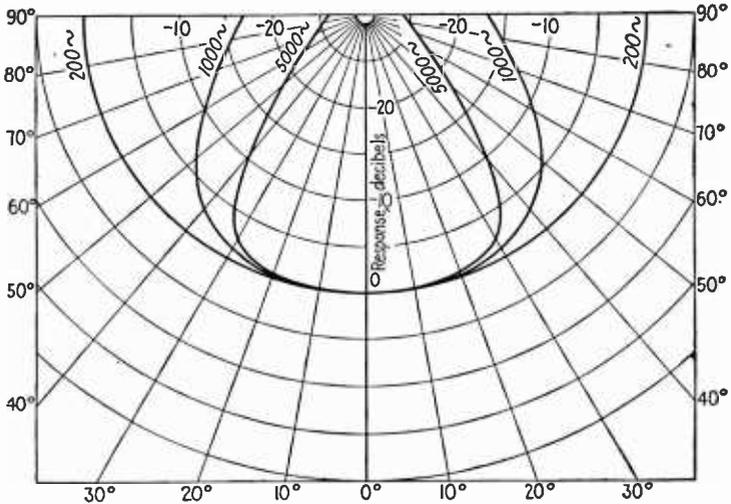


FIG. 21. Axial frequency response of parabolic microphone with a focal length of 8 in.

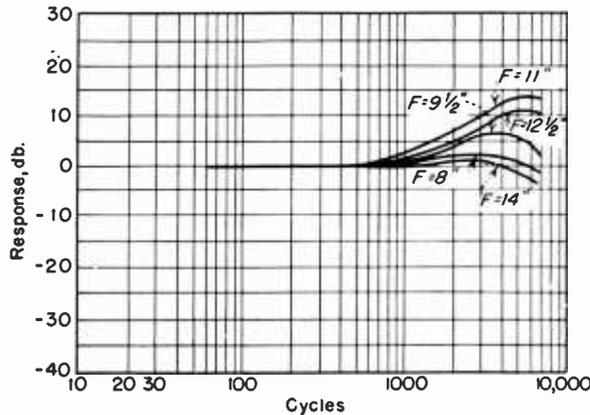


FIG. 22. Frequency response on axis of parabolic reflector microphone at various focal lengths.

because at certain points of focus there is a tendency for sharp irregularities in the l-f response owing to cancellation between the directly received and reflected sound from the paraboloid reflector. In certain instances where the h-f absorption is considerable, the ability to accentuate the highs by refocusing proves very helpful.

Another distinct advantage of the directional microphone is its ability to disregard to quite an extent the acoustics of the room as it responds almost entirely to the sounds upon which it is directly focused. In some cases another microphone without a reflec-

tor has been used with the parabolic microphone so that it may be faded in at certain times to make the reproduction sound more realistic. The parabolic microphone has been used to pick up sound from a certain section of a large crowd or audience of a sports event or to pick up the voice of a single individual at a time in an audience.

14. **Microphone Calibration and Testing.** The sensitivity of a particular microphone is generally expressed as the open-circuit output voltage generated at the microphone terminals for a unit sound pressure against its active element. The intensity of the sound waves impinging upon the active element may be evaluated as a pressure or force. This is usually expressed in dynes per square centimeter.

The actual voltage generated by the microphone being very minute (a very small fraction of 1 volt for a sound pressure of 1 dyne per sq cm against its element), the sensitivity may be expressed in minus decibels below a standard reference level usually taken as 1 volt. When it is desired to correlate this value with the amplifier gain, one would have to assume the microphone to be loaded with a matching impedance. This would result in dropping the output voltage to one-half the open circuit value or a corresponding 6 db decrease in output.

The sound pressure at a particular point where a standard microphone is set up is generally measured by the Rayleigh disk method. This instrument consists of a light circular mirror suspended by a fine quartz fiber at an angle of 45 deg to the axis of the tube through which the sound waves pass. The torque produced on the disk mirror by the sound field is measured by the deflected beam of light focused upon it. For small angles of deflection, the rotation of the disk is proportional to the sound intensity in the tube and consequently to the intensity in the undisturbed field. The actual value of torque may be determined by a torsion head which has a tendency to return the mirror back to its original position.

Where a sound chamber having suitable acoustic properties to prevent reverberation, at the lower frequencies especially, is not available, wave response calibrations are made in open air in a quiet atmosphere. From a standard microphone calibrated in this manner, other instruments may be compared to it for characteristics.

In determining the response characteristic of a diaphragm-type instrument such as a condenser microphone, use has frequently been made of the thermophone method, the thermophone consisting of two strips of gold foil mounted on a plate and fitted into the recess in the front of the microphone, the recess being entirely enclosed and filled with hydrogen. A d.c. upon which is superimposed an a.c. is passed through the foil and causes fluctuations in the temperature of the foil and the gas immediately surrounding it. These fluctuations in temperature cause changes in the pressure on the microphone diaphragm, and the magnitude of the pressure developed on the diaphragm can be computed from the constants of the system. Thermophone calibration is often referred to as a pressure calibration, since it depends entirely upon the actual pressure developed on the diaphragm and hence does not take into account any effects which may occur when the microphone is used for actual pickup purposes. The response obtained by placing the instrument in a sound field of constant pressure is termed a *field calibration*.

The effect of the diffusion of the sound field and the tendency for most acoustic materials to be more absorbent at high frequencies appear to cause the microphone actually to respond more closely to the field calibration rather than to the pressure calibration.

Previous to the use of any microphone in an actual broadcast or rehearsal, it is carefully tested by speaking into it and having another trained individual listen to the quality of the sound reproduced through a high-fidelity amplifying and loud-speaker system. As compared to the results secured from a standard microphone of known high quality, the condition of the microphone under test can be determined.

The outputs of two or more microphones when connected to a mixing circuit should have the same phase relation so that the output of one will add to that of another. If the output of one microphone opposes the output of another, there would be a reduction in total output and corresponding distortion of various degrees. Therefore, all

microphones used for broadcasting are tested for phasing, and each one is connected to its connector plug so that proper polarity exists with respect to other microphones when it is plugged into a socket or jack.

Correct phasing of two or more microphones may be checked by first connecting one microphone to an associated amplifier input and setting the volume control knob to secure a given output reading on the volume indicator while talking into the microphone. The second microphone is then connected in parallel with the first and with the same volume control setting; both microphones are held close together and the same procedure of talking into them is repeated. A decrease in volume from the previous level would indicate that the polarity of one microphone was incorrect, and a reversal of connections on its plug to provide correct phasing with respect to the other microphone is required as indicated by an increase in volume with two microphones properly phased. Phasing of all microphones used should be checked in this manner so that the phasing of each agrees with all others.

STUDIO TECHNIQUE AND MICROPHONE PLACEMENT

15. Studio Problems. A problem of vital concern to a broadcasting system is that of providing favorable acoustic conditions within its studio or auditorium facilities in order that the effects of reverberant sound from the walls of the enclosures may be kept within desirable proportions in comparison to the sound reaching the microphones directly from the source. Of even greater concern are the problems involving correct placement of microphones with respect to the sound sources within the enclosures, to assure faithful reproduction of each voice or musical instrument, their significant overtones, and a pleasant blending of the groups of voices or instruments.

It is, therefore, by virtue of the selection of a microphone which will faithfully transmit all the actual sounds that occur within its range as well as the correct placement of it within a studio or auditorium having suitable acoustic characteristics that high-quality programs can be produced. Under optimum conditions of reproduction a broadcast listener would hear the same acoustic naturalness of the program from his loud-speaker as he would if he were to be transported to a favorable spot in the studio or auditorium where the sounds originating therein would afford a sensation most pleasing for him to hear.

The major considerations involved in proper studio design such as soundproofing, ventilation, optimum dimensions, and suitable acoustical treatment of the walls have been given.¹ At present we shall be concerned only with the problems of microphone placement, assuming that favorable studio and auditorium conditions exist. Normally, these considerations would be as follows: adequate soundproofing that would prevent undesired extraneous noises from entering a given enclosure, and suitable acoustical treatment of the walls and floor to provide equal absorption over a wide frequency range and give the enclosure in itself a uniform frequency characteristic. It is of considerable importance that the frequency characteristic of the studio or enclosure be considered for high-quality transmission because this characteristic is actually superimposed upon that of the microphone under conditions where the reverberant sound received by the microphone is appreciable as compared with that received directly from the source.

16. Single vs. Multiple Microphone Usage. During the first years of broadcasting, it was a usual procedure to use more than one microphone to pick up a program, especially under conditions where the broadcasting group was rather large. This was necessary on account of rather low microphone sensitivity and the inherently high noise level of the carbon microphones used during that period requiring a placement of those instruments sufficiently close to the sound sources to overcome the inherent background noise of these carbon types. The combination of more than one micro-

¹ Hanson and Morris, *loc. cit.* Sivian, L. J., *Bell System Tech. J.*, 10, 108, 1931. Morris, R. M., and G. M. Nixon, *Broadcast Studio Design*, *RCA Rev.*, October, 1936. Nixon, G. M., *RCA Rev.*, January, 1942, p. 259. Content, E. J., and L. Green, *Acoustic Design and Treatment for Speech Broadcast Studios*, *Proc. IRE*, February, 1944, p. 72.

phone for making a pickup has a disadvantage in that the outputs from the several microphones used were not in proper phase relation with respect to the sound sources. This resulted in considerable distortion when the microphone outputs were combined and fed into a common amplifier.

Improvements in microphones to secure higher sensitivity as compared to inherent instrument noise level has resulted in the possibility of using one microphone at a time. The microphone is located at a sufficient distance from the sound sources so that more than one microphone is not necessary to obtain a good acoustic balance from a group. The practice of using more than one microphone at a time has, therefore, been discouraged whenever possible because of the phase distortion in the sound field resulting therefrom. Under conditions where more than one microphone is used at a time, each should be properly phased or polarized with respect to others as mentioned above.

17. Microphone Placement. For diaphragm types of microphones, the directional characteristics at the higher audio frequencies may require the placement of the broadcasting group located in front of the diaphragm to be contained in an area formed by an angle of 30 deg on either side of the microphone axis.

The frequency characteristics of any diaphragm type of microphone are dependent upon the relative positions of the microphone and the sources of sound in front of the diaphragm. When the sounds approach at right angles to the plane of the microphone diaphragm, a uniform response over the desired range might be obtained. But, if the sounds approach from any other point, it will be found in general that the response will fall off with frequency. This characteristic is illustrated by Fig. 19, which indicates how response varies with the angular displacement of the sound source from the microphone axis. It will be noted that there is a high loss at the higher frequencies for high angular displacements from a point in front of the microphone. Since most musical instruments depend for their quality or timbre upon the presence of overtones, it is obvious that, if these overtones are discriminated against, the quality will be changed materially. If, in considering this loss in the higher frequencies with angular displacement, we apply the limitation that the loss at 5,000 cycles shall not be more than 2 db, then Fig. 19 indicates that, in using a single microphone of the diaphragm type, all the musical instruments of a group should be kept within an angle of 30 deg either side of the microphone axis. Present-day microphones of the diaphragm type have uniform frequency response over much wider angles and permit arrangements of the broadcasting groups as shown in Fig. 23 within an angle of 45 deg on either side of the microphone axis.

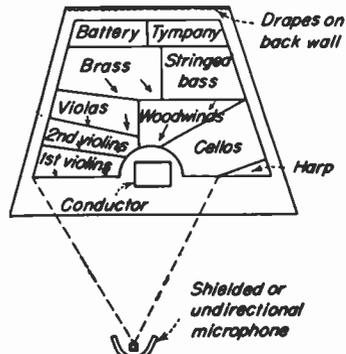


FIG. 23. Setup of 110-piece symphony orchestra.

An individual source of sound such as a speaker, announcer, or musical instrument should not be placed closer to the microphone than 1 ft. Greater distances are determined by the volume range of the voice or instrument and the relative volume desired with respect to the accompanying instruments.

One must consider that in different selections and different arrangements of the same selection the relative importance of the particular instruments may be changed considerably. Where desired prominence cannot be given to a particular group at a certain time using a single microphone, it may be necessary to fade-in another located near the group to be emphasized. A number of microphones can in this way be used, in the particular interval desired, to emphasize the particular musical instrument or instrumental group, the soloist, or the announcer. The control of individual microphone circuits for this arrangement is performed in the control booth by suitable mixing and switching devices. Extensive rehearsing accompanied by listening tests at a

remote point are generally required to secure the desired balance for a particular setup.

In general, the volume range of certain instruments adjacent to one another permits their alliance into natural groups, each instrument of a group being placed approximately equidistant from the microphone. One group may contain violins, violas, and cellos; a second group, the piano, harp, flutes, and clarinets; a third group, the oboes, bassoons, and French horns; a fourth group, the string bass, tuba, timpani, and traps; a fifth group, the trombones and trumpets. In dance orchestras the guitar is usually placed in the first group, the saxophone in the third group, and the banjo with the fourth group.

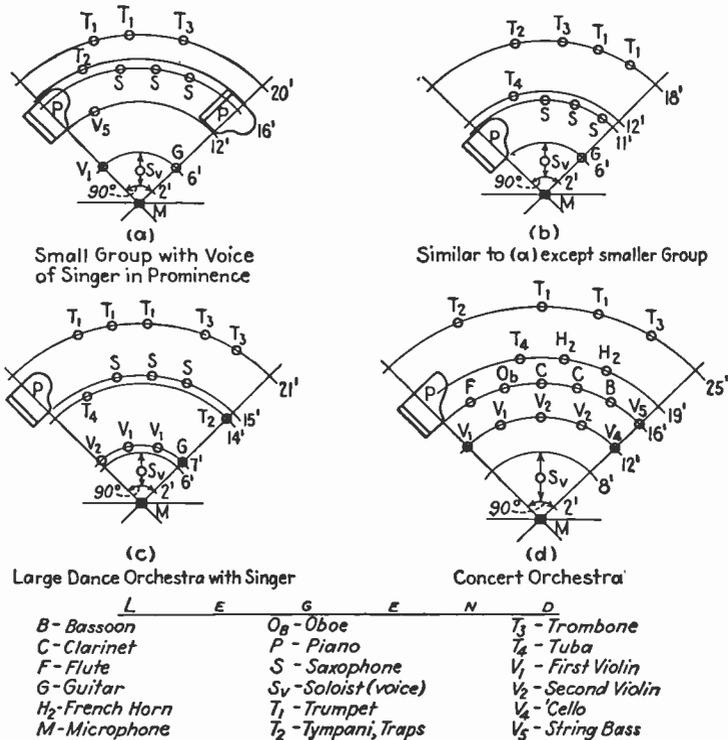


FIG. 24. Orchestra arrangements for use with a single RCA type 88A inductor microphone.

There are many factors involved in securing the proper placement of vocal sound sources or musical instruments before a microphone particularly before a pressure or diaphragm type. While certain rules have been set up, they may serve only as a guide. Most satisfactory results are obtained by a combined study of the instruments as well as an actual setup of them before a microphone in a given enclosure. The results of actual listening tests by means of a high-fidelity speaker and monitoring system performed by one who has a trained ear for music or sound naturalness is a final check upon the proper placement.

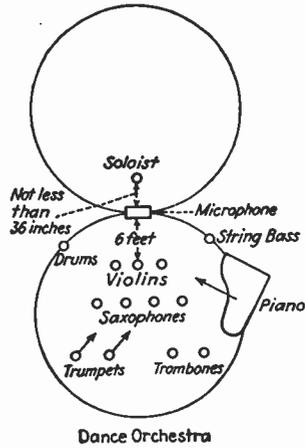
18. Typical Studio Arrangement. A typical setup of a large symphony orchestra before a condenser microphone is shown in Fig. 24.¹ The instruments are placed so as to obtain the desired balance for theater or auditorium work and to obtain the proper harmonic balance allowing for the microphone directional characteristics on higher fre-

¹ Hanson, O. B., *Microphone Technique in Radio Broadcasting*, *J. Acous. Soc. Am.*, 3 (1), July, 1931.

quencies. The microphone is usually placed high above the floor; sometimes near the ceiling of the auditorium.

The microphone is acoustically shielded to prevent reverberation from the auditorium behind it. Present-day microphones, such as the unidirectional and cardioid types, could be used to advantage without the shield since their response in the rear is very small. The string instruments in this setup, being the least powerful ones, are concentrated in the foreground of the group. The wood winds are next in line followed in the background by the powerful brass and percussion instruments. In the arrangement the string tone of the orchestra is given a favorable position to produce a softness to the music which will not be overpowered acoustically by the heavy brasses and percussion instruments.

Figure 23 shows various arrangements of instruments and voices before the pressure or diaphragm type of microphone. The directivity vs. frequency response characteristic of this particular type permits the placement of the musical instruments within an area contained by an angle of 45 deg on either side of the microphone axis. In using this type of instrument the source of sound, speaker, announcer, or musical instrument should not be placed closer than 1 ft from the face of the microphone.



Dance Orchestra
FIG. 25. Dance orchestra microphone arrangement.

The bidirectional characteristics of the velocity microphone are advantageous in that the performers can be distributed on both sides of the instrument in a manner shown in Fig. 25. The uniform frequency-response characteristic of the instrument with directivity is an advantage in that the intensity of some instruments may be

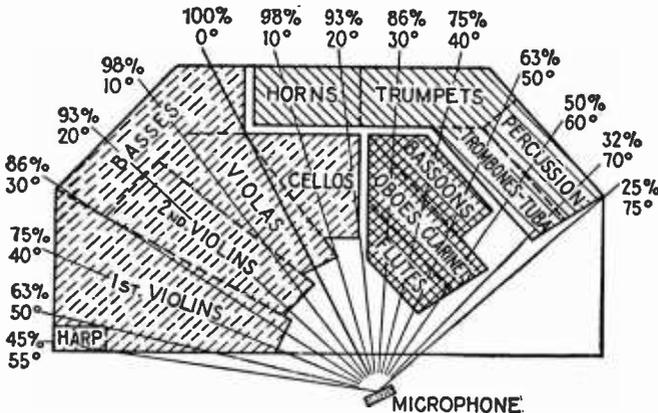


FIG. 26. Velocity-microphone setup for large orchestra group.

decreased without discriminating against their higher frequencies, simply by moving them at a larger angle with respect to the microphone axis.

An orchestral arrangement involving the use of a velocity microphone as suggested by LaPrade¹ is shown in Fig. 26.

¹ LaPrade, Ernest, The Technique of Broadcasting Instrumental Groups, address at North Central Music Educators Conference, Indianapolis, March, 1935. See *Proc. Music Educators Nat. Conf.*, 1935.

The orchestral group in this arrangement was conveniently located on one face of the instrument. To prevent reflection from a wall directly in back of the microphone, the instrument is tilted at an angle of approximately 30 deg toward the orchestra. An exceedingly well-balanced pickup has been accomplished by this method.

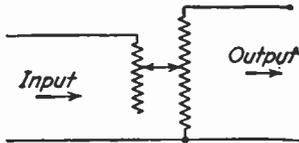


FIG. 27. LT attenuator.

19. Volume Controls or Faders. Volume controls or faders used in high-quality broadcasting circuits should have frequency characteristics which are uniform between 30 and 15,000 cycles to prevent them from causing frequency distortion. Also essential is a very low noise level. This is normally -150 db or better. Proper shielding for protection against dust and dirt is necessary to maintain a low noise level, as well as to act as a shield against any stray r-f electromagnetic fields.

In Figs. 27 to 33 are shown various types of attenuating structures used in broadcasting technique. The type shown in Fig. 27 is frequently used as a microphone fader and is commonly known as the LT structure. When used in multiple such as for mixing several microphone outputs, as in Fig. 28, sufficient resistance is inserted

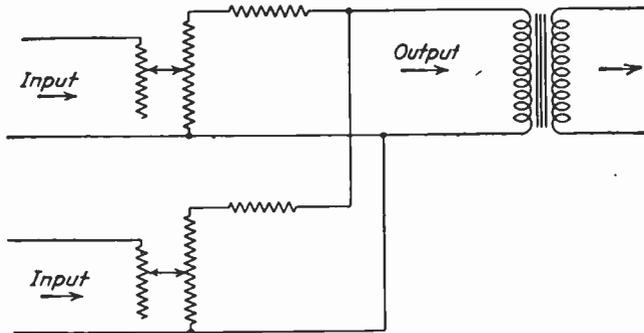


FIG. 28. Multiple-type LT attenuator.

in one output lead from each attenuator to maintain correct circuit matching. The bridged-T structure shown in Fig. 29 is used extensively for the same purposes.

The ladder attenuators maintain an impedance that remains practically constant in both directions through the middle of the attenuation range. Important features of this type of attenuator are its simplicity of design requiring fewer contacts and switches. The minimum attenuation setting of a ladder pad normally corresponds to its insertion loss which amounts to approximately 2.5 db. Where an attenuation range is required extending from zero upward, the H or T structures are used. They are usually constructed with a minimum attenuation setting of zero.

The T and balanced-H structures maintain a constant impedance in both directions when properly terminated. The balanced-H and ladder structures are used where the transmission circuits must be balanced to ground. They are frequently used in broadcasting circuits as master gain controls. Figure 34 shows a high-impedance voltage divider usually in the form of a gain control in the input circuit of a vacuum tube. This is a common type of gain control used on speech amplifier units.

Microphone fading is usually accomplished at high level or at the outputs of the preamplifiers and in a control booth beyond where programs originate in studios. For

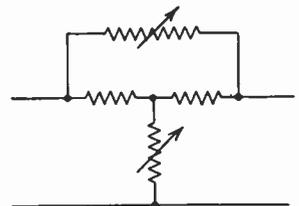


FIG. 29. Bridged-T attenuator.

field pickups the fading is in some cases accomplished directly at the outputs of the microphones. This, of course, requires attenuators of very low noise level. Microphones of the moving-coil dynamic and the velocity-ribbon types have constant low-

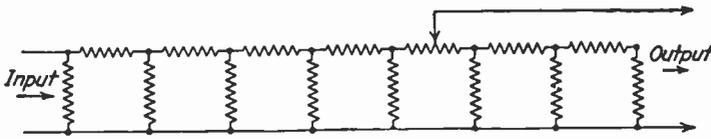


FIG. 30. Single-ladder attenuator.

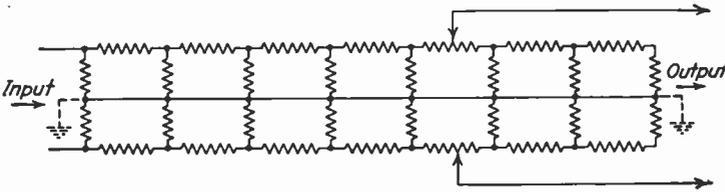


FIG. 31. Balanced-ladder attenuator.

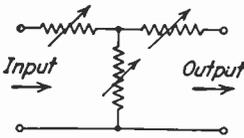


FIG. 32. Type-T attenuator.

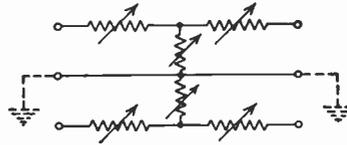


FIG. 33. Balanced-H attenuator.

impedance output over a wide frequency range and for this reason can be faded directly at their outputs.

The curves in Fig. 35 give resistance values of the branches of an H pad suitable for a channel having an impedance of 200, 500, or 600 ohms, the range of attenuation being between 2 and 30 db. Similar curves for other impedances may be determined from the formulas published previously.¹

20. Volume Indicators. The volume level of an audio signal at any particular point in a broadcasting system is normally measured by means of a standardized instrument called the *volume indicator*. The components of the complete instrument consist essentially of a sensitive high-resistance voltmeter of the copper oxide type, an associated T attenuator for extending the range of the meter to higher readings, and a variable resistor accessory to the attenuator provided for calibrating the instrument. The instrument scale is marked in a logarithmic fashion, and superimposed upon this is an associated percentage scale. Two scales are provided, the A type tending to emphasize the vu readings and the B type in which the percentage readings are more prominent.

While an oscillograph placed across the circuits at a particular point in the system would give a true picture of the rather complex wave shapes present from program signals, it would be a rather cumbersome and expensive method of indicating the characteristics of the signal, although it could be used if properly calibrated against a standard. However, through coordination between the broadcasting systems and the

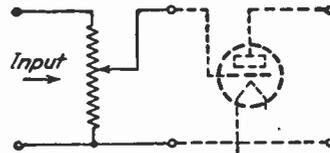


FIG. 34. High-impedance voltage divider.

¹ Johnson, K. S., "Transmission Circuits for Telephone Communication," Van Nostrand; and Lanterman, W. F., The Design of Attenuating Networks, *Electronics*, February, 1931.

telephone company whose facilities were also involved, there was developed¹ a standard instrument of the indicating-needle type having characteristics most suitable for the purpose of indicating signal volume.

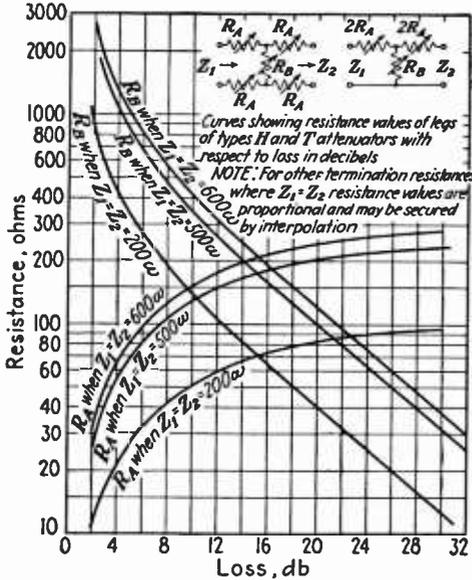


Fig. 35. Chart for H and T attenuator design.

The standard volume indicator is calibrated to read 0 vu when it is connected to a 600-ohm resistance in which is flowing 1 mw of sine wave power at 1,000 cps or n vu when the calibrating power is n db above 1 mw. However, owing to limitations in the present art, it has not been found practical to make an instrument of sufficient sensi-

tivity to be calibrated to read 0 vu across 600 ohms with 1 mw, and therefore the instrument is normally calibrated by the application of 1.228 volts rms (4 db above 1 mw in 600 ohms) to the instrument in series with the proper external resistance to

The standard volume indicator (Fig. 36a and b) utilizes a d-c instrument with a noncorrosive full-wave copper oxide rectifier mounted in its case. Arranged for bridging, as in Fig. 36a, across a line, it has an impedance of about 7,500 ohms measured with sinusoidal voltage. Of this impedance 3,900 ohms is in the meter and about 3,600 ohms is external for the purpose of securing required dynamic characteristics. The dynamic characteristics are such that if a 1,000-cycle sine wave voltage of an amplitude to give a steady reading of 100 on the voltage scale is suddenly applied, the pointer will reach 99 in 0.3 sec and then over-swinging the 100 point by at least 1.0 and not more than 1.5 per cent. The frequency response of the instrument is very good as is indicated by the fact that it does not depart from its 1,000-cycle reading by more than 0.5 db between 25 and 16,000 cps.

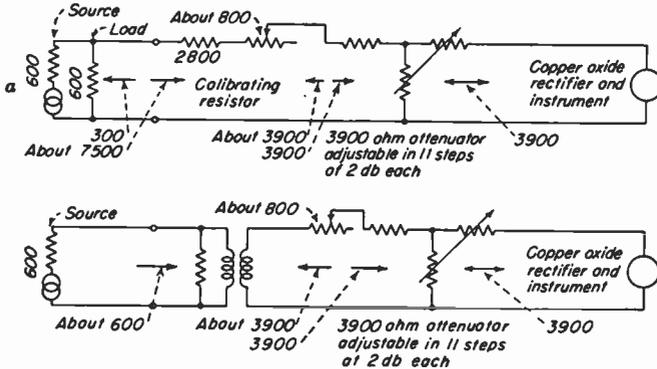


Fig. 36. Standard volume indicator circuit: (a) for bridging across a line; (b) low-impedance arrangement such as line termination.

tivity to be calibrated to read 0 vu across 600 ohms with 1 mw, and therefore the instrument is normally calibrated by the application of 1.228 volts rms (4 db above 1 mw in 600 ohms) to the instrument in series with the proper external resistance to

¹ Chinn, Gannett, and Morris, *loc. cit.*

cause a deflection to the 0 vu or 100 scale point. The instrument therefore has sufficient sensitivity to be read at its normal 0 vu point on a volume level of +4 vu, which is the minimum setting of the attenuator and volume indicator; for this reason the 1-mw calibration is correct.

For routine checking of the calibration of volume indicators, a "reference" instrument properly calibrated may be used in a simple comparison method.

To the terminals of a source of a-c voltage of adjustable output, the reference-volume indicator and volume indicator to be calibrated are connected in parallel. The attenuators of both indicators should be set at +4 vu. The applied voltage is then adjusted until the reference-volume indicator pointer is at the 0 vu or 100 mark. If the pointer of the volume indicator being checked is not then on the 100 mark, its calibration resistor should be adjusted until it reads the same as the reference-volume indicator.

Inasmuch as the standard volume indicator has been developed and standardized as a method of checking volume of signals of complex wave shapes, it has associated with

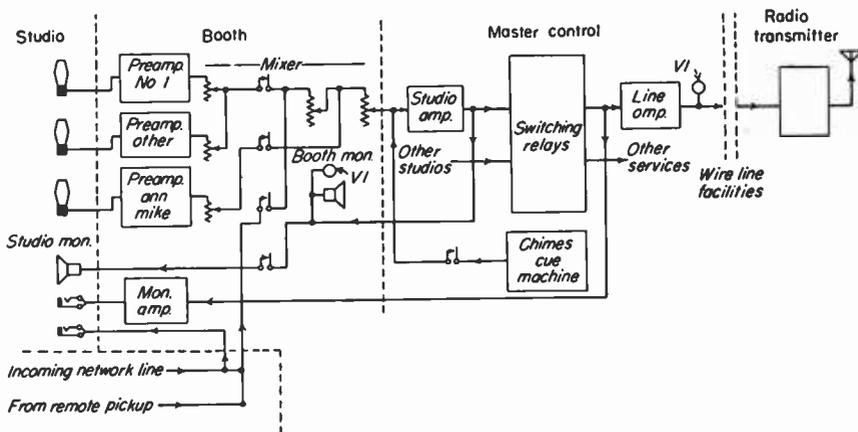


FIG. 37. Simplified diagram of equipment and facilities required for a single program channel.

it the term vu. This term has been restricted to its intended use; hence, whenever a volume level reading is encountered expressed in so many plus or minus vu, it will be understood that the reading was made with an instrument having the characteristics of this standard instrument and expressed with respect to the reference level. Most previous types of volume indicators, even when recalibrated to a 1-mw basis, will not give indications corresponding to those of the new instrument on all types of program waves owing to the particular characteristics of the new instrument.

21. Speech-input Amplifiers. These amplifiers are sometimes termed *preamplifiers* or *microphone, line, and program amplifiers*. They comprise the apparatus necessary to increase the electrical energy output of the microphone or transcription reproduction to a sufficient level to permit its transfer by means of wire lines to the broadcast transmitter. The normal energy level of programs entering the wire lines or program loops is approximately +8 vu (+14 vu delivered from the line amplifier with a 6-db isolating pad). A schematic of the equipment shown in Fig. 37 shows position of the microphone controls, volume indicators, monitoring amplifiers, and relay-switching systems.

Speech-input equipment is designed to have a substantially uniform response from about 30 to 15,000 cycles and above. The maximum gain of such a three-stage amplifier from input to output is approximately 75 db. The input impedances are 62.5, 250, 600 ohms, and the output impedances are 250, 500, and 600 ohms.

PROGRAM RECORDING FACILITIES

The essential parts of a large broadcasting system usually include the facilities for recording programs for the following reasons:

1. To have an accurate record or log of the program material actually broadcasted from a station. This is known as *reference recording*.
2. To secure a record of a studio or special events program at some distance over wire lines and thus be able to reproduce the program at a time most convenient for an audience which may be in time zone a number of hours different from that in which the event takes place.
3. For production of recordings to use at small stations where wire-line facilities are not available.
4. The recording of an audition of a person or group of persons qualifying for a program part.
5. Production of sound effects such as crowd noise, for convenient use and introduction into a particular program.

22. Recording Equipment. The essential equipment required for producing high-fidelity recordings on disk records consists of the following: For bridging a program bus by means of a multiple point switch there is a limiting amplifier of the type similar to that described under Radio Facilities. It is the function of this limiting device to prevent overcutting of the record on high peaks. Following this are duplicate recording channels, each a program amplifier having linear amplitude characteristics and a uniform frequency response over a wide range. This amplifier normally has audio power-handling capabilities up to +46 vu so that high audio peaks are not distorted before reaching the cutter head. There is a standard volume indicator across the line following this equalizer since the cutting head is placed after the equalizer, the output of which drives the cutter head.

23. Methods of Recording. It is possible to secure high-quality recording for broadcasting either by recording sound on disk records on a magnetic wire or tape or on film. Disk records are used most extensively in broadcasting technique.

The principal methods of recording sound on film are more commonly used in sound motion-picture technique at present than for radio broadcasting. These methods include the following: (1) variable density, which may be accomplished by using either

a light valve or glow lamp; (2) variable area, accomplished by using a galvanometer "vibrator"; (3) recording with a Kerr cell; (4) film engraving; and (5) a vibrating ribbon (used abroad).

1. *Variable-density Recording.* The light-valve method uses a light of constant intensity; the ribbons of the valve move in response to a voice current and cause a sound track of variable density to be recorded on the film. When a glow lamp is used to produce a sound track, a light source, whose intensity is varied, is focused on a film through a slit of fixed dimensions. Sound tracks produced by these two methods are similar. Variable-density sound tracks are shown in Fig. 38a and c. The average density of the sound track in this case acts as a "carrier" on which the modulations of the sound waves are recorded in less or greater

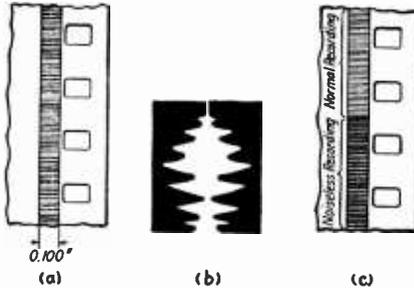


FIG. 38. (a) Variable-density sound track produced by light-valve ribbons or glow lamp; (b) variable-area noiseless track produced by vibrating mirror; (c) noiseless recording showing greater density during periods of low modulation.

density variations than the mean.

2. *Variable-area Recording.* In general this is accomplished by using a light of fixed intensity, which is modulated through the operation of a galvanometer, or vibrator. This produces serrations on the sound-track area of the film, as shown in Fig. 38b.

3. *Recording with Kerr Cell.* In this method the light-valve unit or oscillograph unit is replaced by a Kerr cell. The appearance of the sound track is similar to the variable-density sound track.

4. *Film-engraving System.* In this method an electric-cutting stylus actuated by a power amplifier is used to engrave the sound record directly on the face of the film. The position of the sound track may be inside or outside the sprocket holes. The depth and shape of the groove are similar to those used for cutting disk records, *i.e.*, from 2 to 2.5 mils in depth, and 4 to 6 mils in width.

5. *Vibrating-ribbon Recording.* Several methods developed abroad make use of a vibrating ribbon to cast a fluctuating shadow upon the sound track. One such ribbon valve, developed in Soviet Russia, can be rotated 90 deg, so as to yield at will either variable-area or variable-density recording.¹

The RCA Photophone recorder,² used for variable-area recording, is shown in Fig. 39. Two coils actuate the galvanometer. One carries the voice current to be amplified; the other, a portion of that current which has been rectified and is used as bias. In the absence of modulation a very narrow transparent line is produced down the center of the sound track. A speech signal causes the mirror to vibrate about a central position determined by the bias current and hence to reflect to the film a varying width of the triangular aperture.

A variation of this method is *push-pull recording*, in which the sound track carries two images side by side but 180 deg out of phase. The optical system of the reproducer focuses each recording separately on one cathode of a double-cathode "full-wave" photocell.³

The Western Electric light-valve recorder consists essentially of a duralumin ribbon "hairpin" in a plane at right angles to a strong magnetic field. The ribbon is approximately

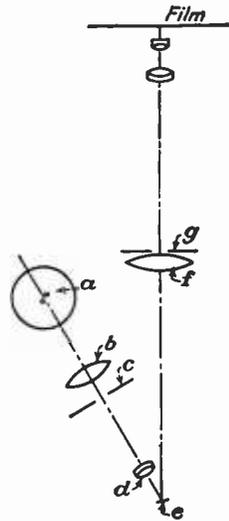


FIG. 39. Schematic diagram of RCA Photophone recorder. (a) Recording lamp; (b) condenser lens; (c) triangular aperture; (d) lens; (e) galvanometer mirror; (f) condenser lens; (g) mechanical slit.

6 mils wide and $\frac{1}{2}$ mil thick. This ribbon is stretched by means of an adjustable spring over a bridge having a narrow slit for passage of the light from the recording lamp through the optical system to the film.

Setscrews are provided to center the ribbon accurately over the slot, which is approximately 8 mils wide and 250 mils long.

The ribbon is tuned after proper spacing on the valve to 9,500 cycles or higher, so that its natural period will be outside

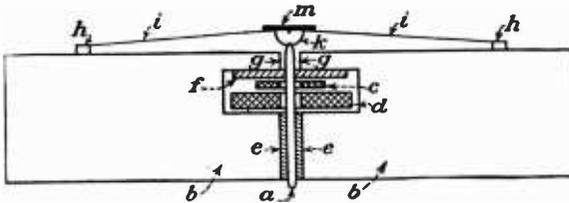


FIG. 40. Schematic diagram of the galvanometer used to actuate the mirror of Fig. 39. (a) Silicon-steel armature; (b,b) silicon-steel pole pieces; (c) voice coil; (d) bias coil; (e,e) non-magnetic spacers; (f) rubber pad for damping at resonance frequency (9,000 cycles); (g,g) air gaps; (h,h) prongs providing tension for galvanometer ribbon; (i,i) galvanometer ribbon; (k) mirror plate; (m) mirror. The mirror vibrates rotationally about a center through the ribbon.

side the range of the frequencies being recorded. A diagram of the optical system

¹ *J. Soc. Motion Picture Engrs.*, March, 1934, p. 158.
² *RCA Rev.*, October, 1936, p. 3.
³ *J. Soc. Motion Picture Engrs.*, July, 1932, p. 51.

using a light valve for recording is shown in Fig. 41. The light source is provided by a special lamp having a horizontal filament. The lamp socket mounting is so adjustable that the filament can be focused properly on the light-valve slit. The sound track produced is shown in Fig. 38a.

A portion of the speech input is detoured through the *noise-reduction amplifier* and used to control a bias current which flows through the hairpin ribbon and in turn controls the ribbon spacing. The result is a *noiseless recording* as shown in the lower half of Fig. 38c. The increase in sound-print

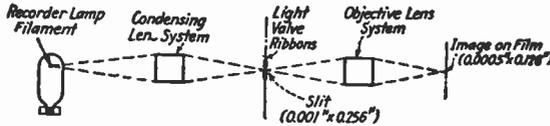


FIG. 41. Optical system used in light-valve recording.

density reduces the ground noise (and consequently increases the volume range of the record) to the extent of about 12 db. A *dialogue equalizer*¹ is sometimes used with wide-range recording to reduce the l-f response during dialogue and especially for intimate close-ups.

24. Glow-lamp Recorder. This consists of a two-element gaseous-discharge tube which varies its illumination in accordance with the voice currents impressed on its circuit. This produces a variable-density sound track similar to the light-valve track. The Aeolight, used by Fox Film Corporation, is one of the recorders in this class. The lamp is not focused upon the film, but a portion of its illumination is allowed to pass through a quartz slit which is in contact with the film.

The recording level for the Aeolight is approximately +12 db above zero reference level. All lamps have a steady d-c component impressed, which causes them to burn at a predetermined exposure. This exposure is modulated by an a-c component due to the introduction of voice currents from the recording amplifier. The resulting output is a variable-density sound track similar to that shown in Fig. 38a. The illumination from a glow lamp is approximately proportional to the amount of current flowing through it, within the normal recording range.

25. Sound on Disk Recording. The direct method of disk recording utilizes aluminum disks usually 16 in. in diameter and 0.050 to 0.060 in. thick, coated with a cellulose nitrate compound (usually mis-called "acetate"). The cellulose nitrate coating is used as the medium for recording sound modulations. These disks are suitable for immediate playback.

For indirect recording it is the usual procedure to use soft wax records approximately 17 in. in diameter and from 1 to 2 in. thick. These records are later processed to produce a hard record approximately 16 in. in diameter and $\frac{1}{4}$ in. thick.

The sound record is cut in the highly polished surface of the wax disk by means of an electromechanical recorder. The technique of cutting wax records is similar to making standard electric phonograph records. The standard speed for common phonograph records is 78 rpm, while for broadcasting records it is usually $33\frac{1}{3}$ rpm. This speed with a 16-in. disk gives a playing time from 10 to 15 min.

Both types of records are usually cut with the spiral proceeding from the outside edge of the record toward the center, similar to making standard electric phonograph records.

26. Variation of Frequency Response on Disks. In recording on a disk revolving at a constant angular velocity, the cutter stylus is placed near the outside edge of the record, and, as the engraved spiral of the sound track progresses toward the center of the disk, the velocity with which the stylus travels on the disk is decreased. This velocity is directly proportional to the radius between the center of the disk and the position of the stylus. Now, for most satisfactory reproduction of the higher frequencies, the stylus must travel with sufficient velocity over the disk to provide sufficient space in the groove to permit satisfactory engraving of the h-f pulsations of very short duration. Therefore there is a tendency for more satisfactory engraving of the

¹ *J. Soc. Motion Picture Engrs.* April, 1934, p. 254.

higher frequencies near the outer edge of the disk than toward the center. In other words, for a given cutter engraving on a disk of constant rotational speed, the frequency response one may reproduce from a disk is more satisfactory near the outer edge of the disk and is less satisfactory, especially to high frequencies, as the stylus moves toward the center. For this reason, if high-fidelity results are to be obtained, the groove containing sound modulation should not be closer to the center of the disk record than 5 in. for 78-rpm recording and 8 in. for $33\frac{1}{3}$ -rpm recording. Medium to good results are obtained with the groove containing the sound modulation at a radius on the disk of not less than $2\frac{1}{2}$ in. for 78-rpm and 4 in. for $33\frac{1}{3}$ -rpm recording. For a given playing time it is sometimes possible to keep this minimum radius, cutting more grooves per inch, sometimes as many as 160 in lateral and vertical disk engraving.

27. Lateral and Vertical Disk Engraving. In the lateral system the groove depth is kept constant, and the engraving stylus moves in a horizontal fashion to produce undulations in the sides of the groove. The groove spacing therefore must be sufficient to prevent the stylus from cutting into adjacent grooves at the low frequencies.

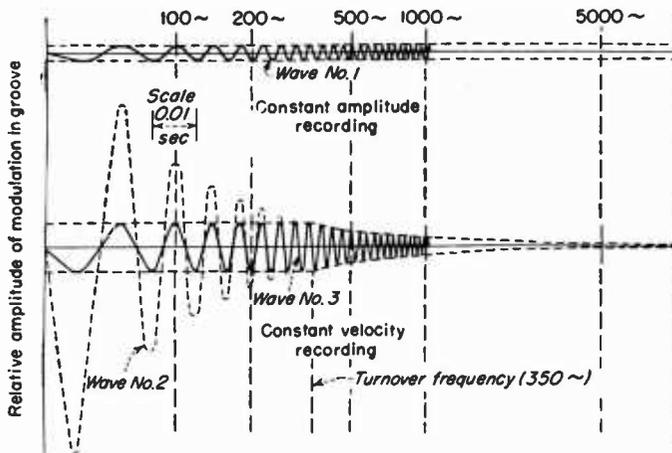


FIG. 42. Characteristic of waves produced with the constant-amplitude and the constant-velocity systems of recording.

The vertical system utilizes an engraving stylus moving in a vertical direction. The groove depth varies with the mechanical modulation, whereas the groove width is kept more or less constant, with a result that the groove spacing can be kept closer with a correspondingly greater duration of playing time.

In Fig. 42 are illustrated waves produced in disk-record grooves under conditions of "constant-amplitude" and "constant-velocity" recording. The wave marked 1 illustrates constant-amplitude engraving produced by a constant sound level regardless of frequency at the cutter. In this case the recorded amplitude is the same for all frequencies.

The constant-velocity system utilizes constant vibrational velocity of the stylus in the record groove under influence of the cutter head. In this case the amplitude of the wave is inversely proportional to the frequency. The wave marked 2 illustrates undulations in the record groove produced by constant-velocity recording, producing an increase in amplitude with a decrease in frequency for a constant sound-level input (assuming that the entire system from the microphone to the cutter head has a uniform frequency characteristic). In this instance, where amplitude = v/kf , the amplitude of a wave frequency f of 100 cycles would be one-half that of 50 cycles for a constant velocity v . At the lowest frequencies, therefore, the amplitudes would be excessive if

sufficient amplitudes of the higher frequencies are to be produced in the record groove. Since the groove spacing would have to be made considerable to avoid "groove cross-over" or "echo" effects in adjacent grooves, due to excessive amplitudes at the lower frequencies, it is customary to cut records at constant amplitude at frequencies below some point between 350 and 800 cycles and constant velocity for frequencies above this point. This is illustrated as wave 3, a solid line. The transition frequency between constant-amplitude and constant-velocity recording, normally some point selected between 350 to 1,000 cycles, is called the *turnover point*.

To produce a constant-amplitude cutting characteristic up to the turnover point and a constant-velocity cutting beyond, it is necessary to utilize corrective equalizers depending upon the particular type of cutter head used. In some cases the response characteristic of the electrodynamic cutter head itself is a contributing factor in the production of the constant-amplitude and constant-velocity range as well as the turnover point. This is illustrated in Fig. 43.

Commercially, direct recording has become of great importance because of its advantages of immediate playback and cheapness when producing disks in small quantities. While the nitrocellulose coating is essentially softer than the pressed records manufactured by the electroplated soft wax process, as many playbacks

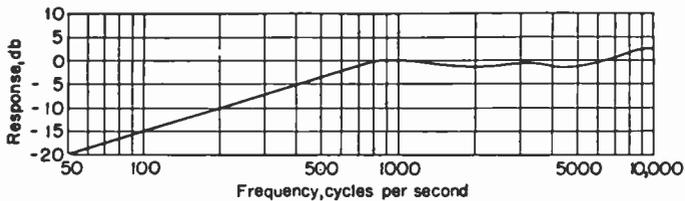


FIG. 43. Frequency response of cutter head based upon optical measurements of the stylus tip motion for constant input.

as 100 may be secured from a nitrocellulose disk with a well-designed lightweight reproducer. A substantially flat frequency response may be recorded on, and reproduced from, these disks over a range of between 50 and 10,000 cycles and higher near the outside portion of the disks. It is good practice where very high-fidelity reproduction is required on 33 $\frac{1}{3}$ -rpm disks to use the outside portion of the disk to compensate for the loss of the higher frequencies in reproducing as the pickup moves toward the center of the disk or to divide time into two or more disks, thus permitting reasonably high linear recording velocity of the cutter stylus. A volume range of approximately 55 db has been obtained from nitrocellulose disks using the lateral system of recording and reproducing. With satisfactory operating conditions over-all distortion of the combined recording and playback operations is less than 5 per cent. This over-all distortion is also a function of engraving velocity, decreasing as the velocity is increased, also decreasing with a decrease in engraved depth.

Flutter is a term used to describe vertical modulation produced in the recording groove due to the bounding of the cutter head at a frequency of approximately 30 cps. It is normally caused by mechanical response of the recording head and its associated supporting-arm mechanism under excitation from building noise and other l-f rumble. In observing reflections from record grooves created by a single source of light, the effects of flutter can be noticed in the form of spokes or long spiral patterns extending from the inside over the recorded surface to the outside. Under a microscope this vertical modulation may be seen as a varying width of the cut groove. Manufacturers supply stabilizers which assist in the elimination of flutter.

When recording on nitrocellulose disks, an air-suction nozzle is provided near the cutter to remove shavings or shreds so that they will not interfere with the engraving process and also to provide for safe disposal of this highly inflammable material. Care must be taken to avoid dust, fingerprints, or grit from entering the engraved surfaces

of the disk. Otherwise there is a tendency for increased noise. It is customary to engrave 120 grooves per inch on these disks, although 96 and 112 and as high as 160 grooves per inch have been used. This number is fixed by the lead screw of the recording machine. The groove depth engraved on this type of disk is normally about 0.0015 to 0.002 in. Commercially, it has been possible to secure recordings of this type having a noise level 50 to 60 db below the maximum modulated signal, although the average record has only a 35- to 40-db spread between noise and modulated signal. By the method explained below for processing soft wax from which pressings are made of a hard material, nitrocellulose disks may be similarly processed for the purpose of making a large number of pressings.

The indirect recording method requires considerably more equipment and time to manufacture the pressed disks than the direct method described above. However, for mass production, pressings can be made considerably more cheaply than single records by the direct process.

28. Necessary Equipment. Equipment necessary for wax disk recording consists essentially of a machine lathe especially designed to turn the wax record clockwise at a uniform speed, which is $33\frac{1}{3}$ rpm for broadcasting work. The carriage of the lathe is driven with a lead screw carefully machined to move the recorder holder at a predetermined rate while cutting the wax record. The lead screw is driven through a gear train which regulates the number of grooves cut per inch, usually 86, 92, 98, 112, or 120. A recorder holder provides the necessary support for the electrical recorder.

A horizontal turntable, driven through a vertical shaft, is provided for supporting the wax record. The vibration of the driving motor is eliminated on different lathes by various methods. The Western Electric lathe uses an oil dashpot placed below the lathe bench, and through which the vertical shaft of the turntable is driven. This dashpot provides the necessary damping to ensure smooth recording on the record.

The RCA machine utilizes a motor on a rubber isolating mounting. The table is driven by means of a rubber roller, the shaft of which is belt driven from the motor pulley.

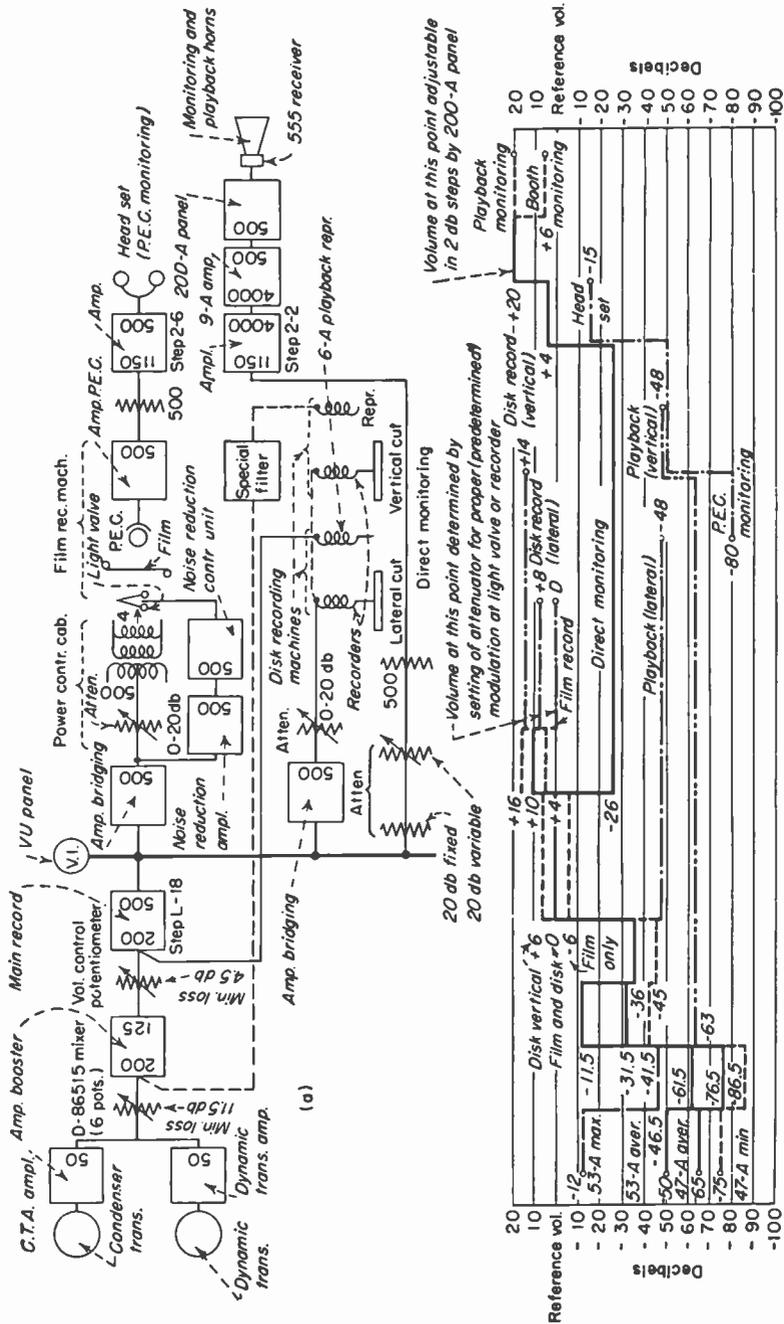
The details given below refer to lateral-cut records, this being the most common type of record that has been used for broadcasting. Vertical-cut records are made by some studios for playback purposes. Both types have their particular advantages.

29. Sound-recording Channel. A schematic diagram of a typical recording setup is shown in Fig. 44 which represents a Western Electric system.

30. Preliminary or Booster Amplifier. This amplifier (see Fig. 44) is mounted between the mixer panel and the volume-control panel. It is used to amplify the output of the mixer before passing through the volume-control panel. Amplification is desired at this point to raise the recording level sufficiently high to prevent undesirable pickup from stray electric currents or other sources entering the voice-transmission circuit. It also eliminates possible noise when operating the volume-control potentiometer. This amplifier differs in detail for various systems. In the Western Electric system, it is a three-stage resistance-coupled amplifier using three 264-A tubes.

31. Volume-control Panel. The outputs from the individual mixer panels are connected in parallel, and leads from them are connected to the input of the preliminary or "booster" amplifier. The output from the preliminary amplifier is fed into a control potentiometer, which permits simultaneous adjustment of the total volume without changing the relative adjustments of individual mixer values. This panel also mounts an extension volume indicator to give a visible indication of the volume level maintained at the bridging bus.

32. Main Amplifier. This amplifier is so designated that it amplifies the output from the volume-control potentiometer and delivers the amplified current to the bridging bus circuit (or in simpler installations, directly to the power-control panel and recording machine). It is the amplifier furnishing the largest gain in the recording channel. The main amplifier differs in details for the several recording systems. In the Western Electric system it may be an impedance-coupled amplifier with input and output transformers, *i.e.*, the first stage using a Western Electric 102-type tube and the



Figs. 44a and b. (a) Schematic of Western Electric recording system; (b) transmission-level diagram of the system.

second and third stages, 205-type tubes. The total gain of this amplifier is approximately 70 db. The gain control of the amplifier is provided by a potentiometer in the input circuit. One bridging amplifier is required for each recording machine, its principal function being to prevent variation in individual recording circuits from introducing any loss or distortion to other circuits. It divides the electrical circuit output from the main amplifier, depending upon the number of amplifiers connected to the bridging bus. It is essentially a power amplifier, with the input transformer arranged for a high input impedance, making the bridging of several of the amplifiers across the main bus practical.

The bridging-amplifier outputs are connected to the film and wax recording machines in the recording room. The wax recorder requires approximately +8-db volume level, and the film recorder around 0 db.

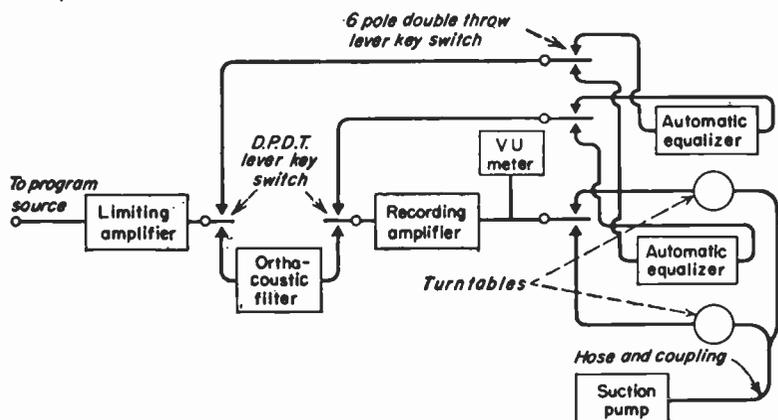


FIG. 44c. Recording equipment layout for two type 73-B recorders.

33. Disk Records. The grooves of a disk record are ordinarily spaced 92 to 160 per inch. For 92 grooves per inch this allows about 0.011 in. from center to center of the groove, of which 0.006 in. is the width of the groove itself. The maximum lateral motion of the stylus is thus limited to about 0.0025 in. on either side. Generally, 0.002 in. should not be exceeded. Cutters usually used are designed as constant-velocity devices. In practice such cutters have this characteristic only above 300 cycles or higher. Below this point the amplitude is independent of frequency. If the maximum amplitude for a 300-cycle wave is equal to 0.002 in. on either side of the center, then a 1,500-cycle amplitude for the same electrical input level would be 0.0004 in.

The shape of the groove varies somewhat in commercial practice, but it is approximately 0.006 in. wide and 0.0025 in. deep. The pitch of the groove is generally 0.010 to 0.011 in., leaving a space between grooves of about 0.004 in. With only this space available, the maximum safe amplitude is something less than 0.0025 in., if the walls of the groove are not to be cut too thin.

Cutting stylus consists of a sapphire, synthetic ruby, or other hard point fastened to the lower end of the stylus arm. One end of the sapphire has a rounded point about 0.002 in. radius and standard cutting angles of approximately 70 or 90 deg for the sides.

The advance ball is a small cylindrical sapphire, ground spherically at one end and held in an adjustable mounting attachment to the recorder. This ball supports the weight of the recorder, and the arm, being adjustable, permits regulation of the depth of the groove on the wax.

Playback reproducer is provided to permit playing back the wax record immediately after it is cut for rehearsal work and test. This usually renders the wax unsuitable

for processing, and for this reason two wax records are usually provided for each recording channel, one of which can thus be used for playback and the other for processing. The pressure of the needle on the wax is generally adjusted to between 15 and 20 g.

A needle provided for playback from the soft wax is designed differently from the ordinary needle used for the finished hard record. The Western Electric type has a point of 0.003 in. radius. The needle is constructed on a mandrel, ground to a smooth finish, and the point given a chromium plate to improve wearing quality.

Checking Speed. The periphery of the turntable is usually divided with vertical lines, so that a neon lamp, operating from a 60-cycle source, may be used as a stroboscope to observe the turntable motion. The lines on a standard turntable are usually arranged so that with 60 cycles on the lamp, as the turntable rotates at exactly $33\frac{1}{3}$ rpm, the lines will appear to be stationary. If faster than $33\frac{1}{3}$ rpm, the lines will advance slowly, and, if slower than $33\frac{1}{3}$ rpm, the reverse will be the case. This check of the speed is usually made with the wax record on the turntable.

Checking the Damping Action. A method of checking the instantaneous constant speed may also be used to check correct damping of the turntable. With the turntable rotating at normal speed, the oscillator for supplying 60-cycle source to the neon lamp may be adjusted until the vertical lines appear stationary. If the disk is now touched lightly by hand, the line or spot observed will appear to shift its position owing to momentary load. As soon as the hand is removed, the line or spot observed should come back to its original position. Observing the movement will determine whether the turntable has insufficient damping or too much damping.

Determining the Starting Point. Disk records for radio broadcasting use are usually cut in clockwise rotation from the outside in, similar to ordinary phonograph disk records. To obtain a definite starting point for the records when in use, the first groove is spaced an appreciable distance from the rest of the cut. This is obtained by a coarse speed cam actuating the lead screw at the start of recording. As the lead screw makes its first complete revolution, it moves the recorder under the influence of the cam until the recorder is in its normal cutting position.

34. Wax-suction Equipment. This equipment is provided to furnish a means of removing the shavings from the wax record during recording. The suction tube is so placed that the shavings thrown off by the stylus are carried away from the face of the wax. A central suction system is usually provided in studios having several recording channels. This usually consists of a turbine suction pump with pipe lines leading from a central suction point to a separator tank placed in each recording room. In some smaller installations, an individual bell jar, with a small suction motor, is used for each recording machine.

35. Wax Preparation. Two types of waxes are generally used in sound recording, those having a working temperature of 75°F, and those with a working temperature about 90°F. Matthews type M, 75°F working temperature, is perhaps most commonly used. It is considered good practice to maintain the room temperature for the type M wax around 75°F when recording.

The procedure for preparing the wax consists briefly of the following steps:

1. At the center of the wax, which is usually indicated by a cross mark, a $\frac{9}{32}$ -in. hole is drilled to a depth of $\frac{1}{2}$ in.
2. A coarse cut is made for a depth of about $\frac{1}{8}$ in. on one face of the wax and repeated as necessary to obtain a perfectly flat surface. The wax is later reversed, the first cut surface becoming the base for the finished wax.
3. On reversing the wax, a hole is cut from the other side to meet the hole drilled on the bottom.
4. A coarse cut is now made on the top surface and repeated where necessary to produce a smooth and flat surface. The wax is now ready for the final shaving or polishing cut, which is done with a sapphire or ruby cutting tool.
5. The face of the shaving knife is usually set at an angle of between 40 and 50 deg to its line of travel, depending upon the particular design of the knife. Its rounded end is toward

the center of the wax. The cutting face of the knife is set at an angle of 90 deg to the surface of the wax. The turntable revolves in a counterclockwise direction.

6. The suction nozzle is placed close to the cutting knife, about $\frac{1}{8}$ in. from the front face and $\frac{1}{32}$ in. above the cutting edge.

7. The best finishing speed is usually determined by experience, but generally ranges from 150 to 160 rpm. The finished cut on the wax should give a perfectly polished surface free from ripples or blemishes of any kind.

36. Record Processing. Briefly, this consists of the various steps after obtaining the soft wax record, to produce the final hard record for commercial use. A complete description of each step would go beyond the limits of this section. The following are the essential steps in this process:

1. The surface of the engraved soft wax disk is rendered conductive by spreading a very thin, extremely fine conducting powder, such as metallic powder, over its surface; by the finer processes of depositing silver from a solution of silver nitrate; or by sputtering pure gold of very minute thickness on the surface. This metal coating is for the purpose of forming one electrode in the electroplating process.

2. Electroplating of this record with a sheet of copper $\frac{1}{32}$ to $\frac{1}{16}$ in. in thickness deposited on the wax. The negative electroplate obtained is separated from the wax and used to hot-press a molding compound, such as shellac, mixed with a finely ground filler. The first electroplate obtained is called a *master*.

3. Two test pressings are made from the first master, after which it is electroplated with a positive.

4. From this positive, sometimes referred to as an *original*, a metal mold or *stamper* record is made.

5. From the record, duplicate originals may be made and, from them, duplicate molds or stampers. By thus making a number of duplicates, it is possible to protect the original master from injury.

6. From each stamper it is possible to obtain as many as 1,000 finished pressings.

Generally, it may be said that the duplicating process reproduces everything on the original wax engraving to such a fine degree that the only difference one may observe is in the materials, one soft wax, and the other a harder, more durable plastic, composed of shellac, vinyl, or acetate compounds mixed into a filler having very little abrasive properties. The surface of these manufactured records is considerably harder than the nitrocellulose coating on metal-covered disks used for direct playback and, with a sufficiently light reproducer, will reproduce with good quality up to 1,000 playings.

37. Re-recording. It is common practice to select desired portions of a sound record by a process of re-recording. This is done with both disk and film records. Either can be played on standard reproducing equipment, which then serves as the input to the recording system, in place of the microphones. Special re-recording equipment is also used; one type consisting of a film reproducer and a film recorder combined in a single instrument and actuated by a single motor. The output of the reproducer photocell is, of course, returned to the recorder light valve in the same casing only after it has passed through an external amplifier. This instrument is used to copy on 16-mm film a sound track that was originally recorded on 35-mm stock; optical reduction, however, is also used for that purpose. Duplication of records and films is often called "dubbing."

Re-recording is used to superimpose special sound "effects" upon a record. For this purpose two or more reproducing systems are connected as a parallel input to the recorder amplifier. The method offers superior control over the relative volume of such sounds as gunshots, background music, or storms and, moreover, tends to reduce the cost of production. A library of "effect" records is maintained at many studios.

Originals intended for re-recording are sometimes made abroad by cutting a lateral track in discarded film, which is reported to be entirely serviceable for this purpose and to withstand many playbacks without damage.

38. Electrical Recording Machines. It is essential that a recording machine of a precision type should have a constant speed. For this reason it is usually driven by a

synchronous motor. The mechanical inertia of the revolving table assists in keeping the rotational speed constant, the speed regulation of the disk being usually better than 0.3 per cent. It is customary to mount the driving motor on vibration dampers in such a manner as to prevent the motor vibration from reaching the revolving table. Vibration from the motor shaft is kept from reaching the shaft of the revolving table either by using belt drive, rubber differential speed rollers, or both. The spacing of the grooves cut on the disk is controlled by gear trains and the lead screw which moves the cutter head toward the center of the disk. The number of grooves engraved per inch can be set by means of the gears. A suction tube is provided for removing the shaving or thread produced while engraving. A microscope and groove illumination lamp facilitate examination of the engraved grooves. A playback pickup arm is generally provided in addition to the engraving cutter mechanism to permit playback of the record for quality checking.

39. Recording Heads or Cutters. The essential requirements of a recording head suitable for producing high-quality recordings are as follows: (1) freedom from amplitude distortion in producing undulations on the disk record, (2) suitable frequency-response characteristic over a range of 40 to 10,000 cycles to produce constant-

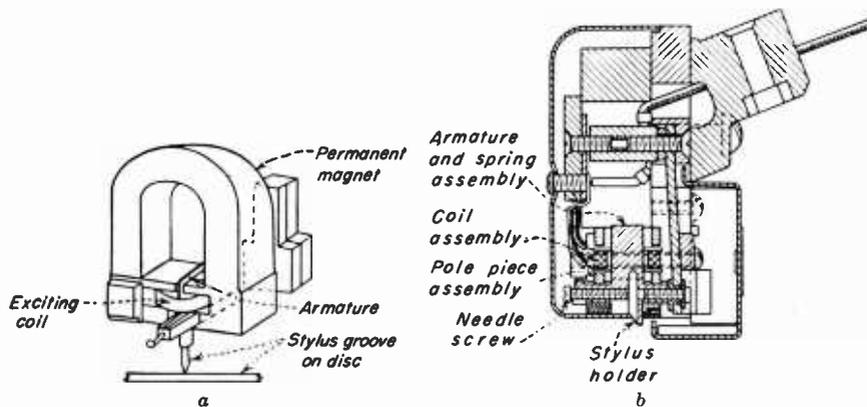


FIG. 45. (a) Electrodynamic type of recording reproducer; (b) high-fidelity recording-head assembly.

amplitude and constant-velocity recording over the frequency ranges required, (3) freedom from mechanical resonance which would tend to cause overcutting, and (4) reasonably good efficiency in transformation of complex electric wave energy into mechanical vibration of the cutting stylus.

There are numerous types and designs of cutting heads manufactured for recording sound on disk. The most common in present-day usage are the electrodynamic and the piezoelectric crystal types.

Electrical recorder heads provided for disk recording are generally designed so that the average linear velocity of the stylus (which may be expressed as a constant \times the frequency \times amplitude) is proportional, over a wide range of frequencies, to the impressed voltage, or $v = kfe$. The method of damping the moving system varies with different records. The Western Electric recorder uses a rubber tube about $\frac{1}{2}$ in. in diameter and 8 in. long, one end of which is fitted to the armature assembly and the other end free. Oil is sometimes used to damp the armature movement in other types of recorders.

A drawing of an electrodynamic type of recording cutter is shown in Fig. 45a. With a modulated current passing through the winding of this instrument, the armature produces and transfers to the cutting stylus mechanical undulations conforming with those in the electric wave, except that the amplitude is altered somewhat by

mechanical and electrical means. In Fig. 45*b* is illustrated the RCA MI-4887 high-fidelity recording head. This cutter head utilizes a band-pass mechanical network terminated in a dry mechanical resistance material. The balanced armature is centered by means of a tempered steel spring. It is supported on knife-edge bearings upon which the lateral stylus motion is centered. Nicaloi is used for the pole pieces of the permanent magnet.

The frequency-response characteristic of this cutter head is shown in Fig. 43. Below 800 cycles, frequencies are controlled to hold amplitude constant, the stylus velocity decreasing as the frequency is reduced. Above the 800-cycle point the response curve shows constant-velocity motion well over a frequency of 10,000 cycles. It is possible by electrical means to move the turnover point in this curve from 800 cycles to a lower frequency of, say, 500 cycles if desired.

While the electrical input impedance of the cutter head itself is approximately 5 ohms, an electrical impedance compensating network can be secured to retain the total input impedance of 15 ohms throughout a wide frequency range. A high-quality amplifier having a power output of 10 watts or more is recommended for driving this cutter head.

40. Crystal Cutting Head. This type of recording head utilizes a 5-ply bimorph Rochelle-salt crystal to drive the sapphire stylus to engrave sound waves laterally on disk records. For the constant-amplitude recording range the voltage applied to the crystal of the cutter head is normally 75 volts rms, while for the constant velocity range of recording it is about 150 volts rms. Since the internal impedance of the head is rather high normally, 159,000 ohms at 100 cycles, the actual power consumed by the crystal is rather small, being less than 1 watt, although the power output recommended from the driving amplifier is considerably more.

A corrective equalizer is required with the cutter for constant-velocity recording above 350 cycles. Under correct operating conditions the manufacturer shows that this cutter has a frequency characteristic substantially flat within ± 3 db between 30 and 10,000 cycles.

A sapphire cutting stylus is recommended for use with the cutter head. For most conditions of recording the groove depth is 0.0025 in. for cutting soft wax and 0.0015 to 0.002 in. for nitrocellulose records.

41. Measurement of Frequency Response. By examination of the frequency-response curves of the various component parts of a recording system the over-all performance of the system can be checked. The program microphones and amplifiers which feed the recording head are measured in a conventional manner with a standard sound source, beat-frequency oscillator, output meter, or cathode-ray oscillograph. Under these conditions the output of the amplifier at the terminals of the cutting head is usually flat within ± 1 db between frequencies of 40 to 10,000 cps.

The recorder cutting head, however, usually has a sloping frequency characteristic (Fig. 43). The response of the cutting head alone has been measured by supplying constant level tone at various frequencies to the head and, by means of a tiny mirror attached to the stylus, reflecting a beam of light into a phototube. It is usual practice to measure the response of the cutter and disk material together.

This consists of making a recording of the output of a beat-frequency oscillator held at constant voltage at the cutter terminals. Frequencies usually recorded in order from outside to inside are as follows: 10,000, 9,000, 8,000, 7,000, 6,000, 5,000, 4,000, 3,000, 2,000, 1,500, 1,000, 800, 500, 300, 200, 150, 100, 80, and 50 cps. The completed record is then removed from the turntable; and under a concentrated single source of light, the reflection of light source as seen in the grooves shows peculiarly patterned shapes similar to their descriptive name "Christmas tree." The pattern is symmetrical about the radius of the disk. It is actually a graphic representation of the frequency responses of the cutter and disk material together. The radius of the disk is the axis of frequency, the end of the pattern nearest the center being the lowest frequencies. The width of the pattern measured perpendicular to the disk radius is proportional to the undulations of the groove. This in lateral recording corresponds

exactly to modulation depth. This phenomenon is due to the reflection of light over a wider band, the greater the ratio of modulated groove width to depth.

Inasmuch as good reproducing equipment usually has flat characteristics, the Christmas tree pattern may be produced with straight sides from the turnover frequency, of say, 500 to 7,000 cycles. Below this, it is customary to compensate the loss of low frequencies by boosting them with electrical filters in the reproducer. If it is noticed that pronounced peaks are in the pattern, the cutter head may be adjusted or filters inserted to produce the response characteristics required.

42. Record Reproducing Facilities. Transcribed programs generally originate in studios located separately from those in which recording is done. It is quite evident that, if full advantage is to be taken of the high-quality program material recorded on disk records, the transcription or reproducing equipment must also be of the precision type.

The transcription turntable is generally driven by a high-torque synchronous motor cushion-mounted within the console or cabinet. The motor shaft is flexibly coupled to the main turntable spindle. Speed regulation is reduced to a very small value for both rotational speeds of $33\frac{1}{3}$ and 78 rpm by means of flywheel inertia and a mechanical filter on the drive shaft.

Speed reduction of the RCA type 70-C1 turntable is accomplished by means of a heavy-duty ball-bearing speed-reduction mechanism operated by a button located at the rim of the turntable disk. Noise and vibration pickup is kept at a minimum by cushion-mounting the motor and spindle housing and cushioning the suspension arms.

Special consideration is generally given to the design of a satisfactory tone arm and reproducer head for high-fidelity reproduction. The reproducer head must be light in weight and in pressure on the groove of the disk. Normally the pressure exerted by the diamond point stylus as measured by means of a spring balance or postal scale should not exceed 2 oz. A more desirable weight is less than $1\frac{1}{2}$ oz. A lightweight tone arm and reproducer head assists in the reduction of record hiss or scratch noise and also the reduction of high frequencies especially near the center of the disk. Lightness also assists in securing more playbacks from a recording, since a lateral reproducer having a stylus which operates too stiffly or having too great a pressure on the disk tends to erase the higher frequencies from the record groove.

Commercial reproducer heads generally utilize electrodynamic or piezoelectric principles as electric generators to convert mechanical force supplied by the groove modulation through the stylus assembly to the electrical generator element.

The RCA MI-4856 reproducer (Fig. 45a) is equipped with a permanent diamond point, the radius of which corresponds to the 0.0023-in. standard for lateral cut non-abrasive high-fidelity records. The armature is of the clamped-reed type. The two upper air gaps are filled with nonmagnetic material and are inactive. A linkage having

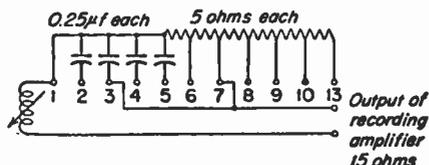


FIG. 46. Circuit for correcting characteristic of recording head.

6:1 leverage ratio is provided since the armature impedance is too high to be directly coupled to the record groove through the stylus. A diamond point is secured in the lower end of an extremely light pivot-arm spring supported vertically but rigid laterally. The pivot arm is thus permitted to rise without lifting the entire head. In the direction of useful motion transmitted to the armature the linkage has a minimum of compliance with a resultant cutoff of about 9,000 cps. This peak is reduced by means of a block of loaded rubber arranged as a selective damper approximately adjusted for the resonant frequency.

A shunt capacitance located within the tone arm is generally connected across the pickup coil to react broadly with the inductance, increasing the response through the upper frequency range. An equalizer may be placed directly at the output of the pickup head to compensate for losses in the record modulations.

The piezoelectric type of lateral disk-record pickup head utilizes a bimorph crystal under torsional strain to convert mechanical modulations of the record groove into electrical waves. The sapphire stylus used with this reproducer is set in a small screw which fits the thread of a hollow magnesium block. The motion of the chuck is converted into a torsional strain in a bronze wire. This in turn conveys a twisting force to the bimorph crystal sealed hermetically within a compartment. The emf produced at the electrodes of the crystal is developed from the twisting force produced by the stylus and attachment mechanism.

This type of reproducer head is normally rather light in weight, resulting in a stylus pressure of approximately 1 oz on the disk. It may be used for reproducing either constant-amplitude or constant-velocity recordings, the type of electrical compensating network required being dependent upon the particular characteristics of the recordings.

43. Orthacoustic System. There is a limitation in the amplitude of the lower frequencies recorded upon a disk. This is corrected by a sloping characteristic in the response curve below the turnover point, brought about either mechanically in the recording head or electrically with a suitable network. The undesirable needle h-f hiss is another limitation to be overcome for a satisfactory recording and reproducing system, as well as the l-f rumble caused by the turntable and building vibration, etc.

In the RCA-NBC Orthacoustic system, recording and reproducing units are individually compensated to offset characteristics of each other and thus create a reproduction which is very nearly the equivalent of the original sound. Below 100 cycles the characteristic of the recorder system is made constant velocity by electric means. This tends to give preemphasis to the low frequencies. Then it rises from 100 to 500 cycles on a constant-amplitude basis in accordance with the mechanical and electrical characteristics of the cutter.

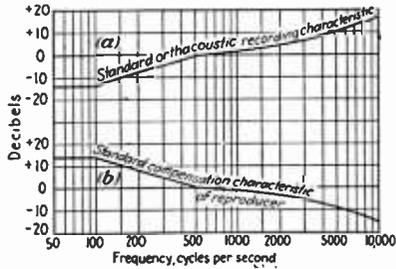


FIG. 47. Orthacoustic recording characteristic which gives preemphasis to high frequencies.

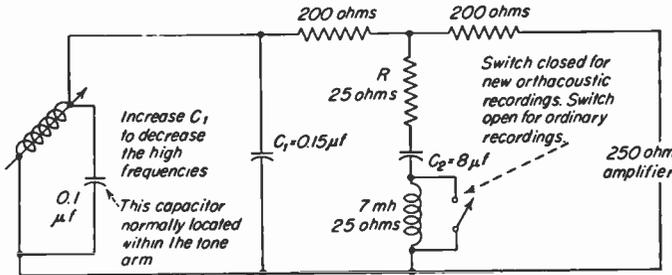


FIG. 48. Compensation filter for orthacoustic reproducer.

Above 500 cycles a preemphasis above a constant velocity is given to the high frequencies, especially over the noise frequency range.

The necessary characteristic for reproduction is the inverse of this curve, or Fig. 47, secured by electrical and mechanical means, especially those of the transcription head itself (see Reproducer or Playback System). An over-all response curve is produced which is flat over the desired range.

44. Magnetic Wire and Tape Recorders. The magnetic system of recording sound patterns upon a record of steel wire or tape by magnetization is illustrated in a simple manner in Fig. 49. The steel wire of small diameter, carried upon required reels,

moves rapidly past the pole piece of an electromagnet while the windings are energized by the audio signal source. In traversing the windings of the sound-head magnet, the audio signal currents of complex wave shape cause a fluctuating field to exist in the pole piece. Corresponding longitudinal magnetic patterns are impressed in the moving wire.

Reproduction of the audio signal impressed with magnetism on the wire is accomplished by connecting the sound-head electromagnet to a telephone receiver or other electroacoustic translating device and running the magnetized wire past the pole piece in the same direction and at the same rate of speed. Under these conditions the magnetic patterns in the passing wire create fluctuations in the magnetic field of the pole

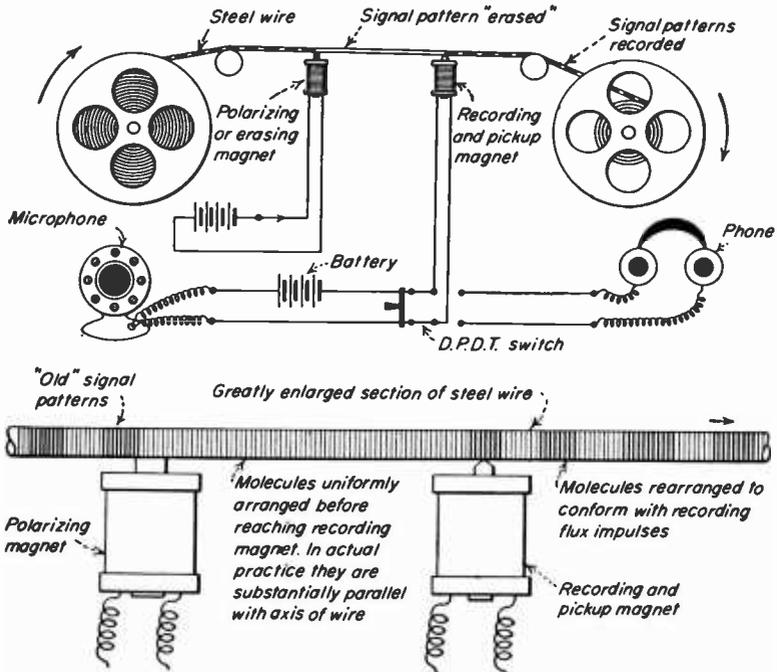


FIG. 49. Essential principles and parts of the magnetic recorder.

piece and generate currents in the windings which correspond to the wave form of the original audio signal.

Recording sound on a magnetic wire offers certain advantages over other methods especially in applications where records of a temporary nature are desired. In this case, the same medium may be used over and over again to make new recordings when old ones are no longer required. The sound record on magnetic wire or tape may be erased and it may again be made ready for reuse by simply causing it to pass through a steady magnetic field which causes the old signal to be erased an instant before the new signal is impressed.

Another advantage is in making continuous long-duration recordings upon a single magnetic wire or tape medium, thus avoiding the procedure of having to change recording disks at intervals when using the sound-on-disk system. A sound record on magnetic wire or tape may be stored away for an indefinite time without fear of deterioration. The wire or tape record may be played back hundreds of times without

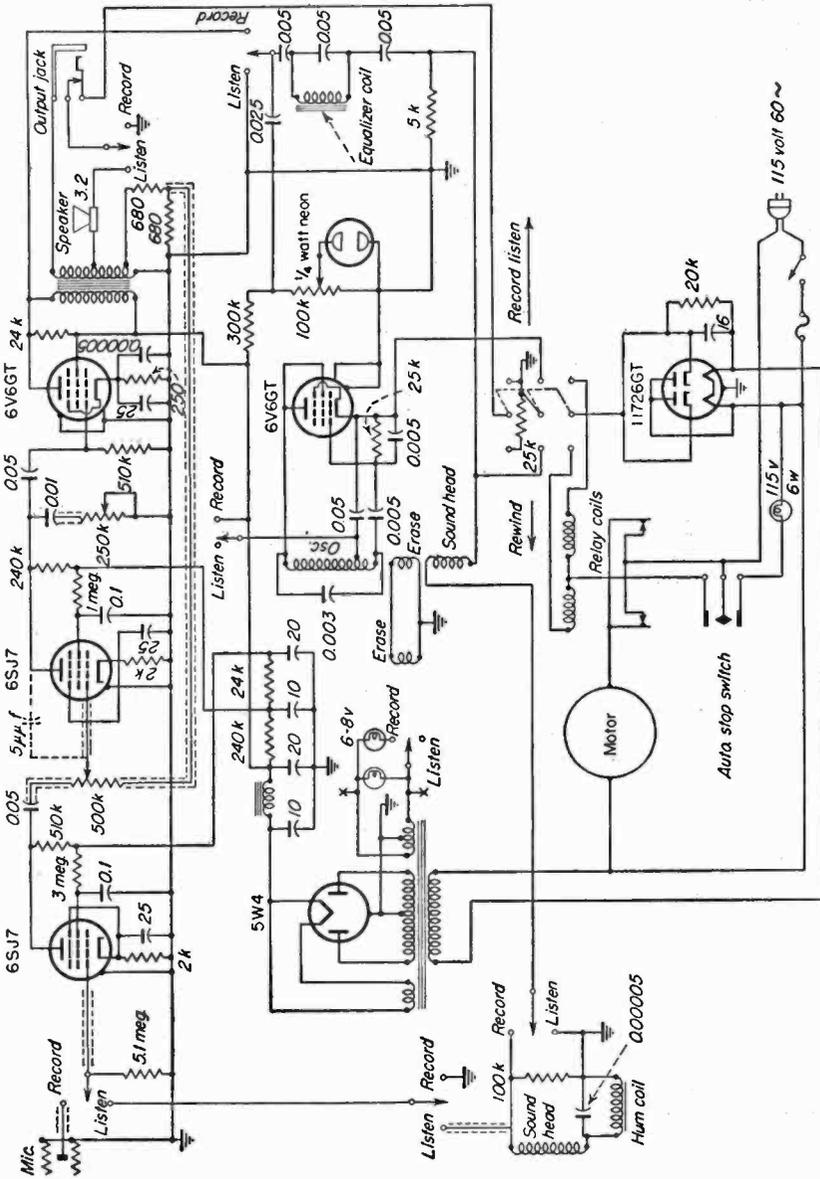


FIG. 50. Schematic diagram of Pierce magnetic wire recorder and reproducer.

appreciable loss of quality which is an advantage over disk recordings which depreciate rapidly when used continuously.

A problem of great importance in recording sound on magnetic wire or tape is the high speed required for the wire to pass the sound head to record the a-f range for satisfactory program reproduction. For example, a speed of 10 fps has been found necessary for satisfactory recording on the wire of frequencies as high as 2,000 cycles, which would mean correspondingly a speed of 30 fps to extend the range to 6,000 cycles. Under these conditions an enormous amount of wire or tape would be required to record 1 hr of program material.

In the case of magnetic wire recording, high wire speed under these conditions is sometimes due to an excessive longitudinal part of the wire required to register satisfactorily without distortion each individual cycle of the program wave. Owing to the tendency of the flux surrounding the pole piece of the recording magnet to spread and influence a larger part of the wire than that directly under the magnetic pole piece at a given instant, the wire speed must be sufficient to permit spacing between the individual cycles of the audio signal.

The Pierce model 55A magnetic wire recorder employs a stainless-steel wire 0.004 in. in diameter as the recording medium. The wire is contained on two spools or drums having equal diameters of approximately 3 in. A constant-speed motor drives the spools forward at a speed of 200 rpm during the recording and reproducing process and at 400 rpm during the reverse rewind process. Pulleys guide the wire between the pole pieces of the sound head which serves both for recording and for reproducing. Figure 50 is a schematic of the circuits of this instrument including amplifiers.

45. Wire Lines. Wire telephone systems are employed almost exclusively for the national distribution¹ of programs to the various stations connected on a network.

The frequency band which is transmitted over long-distance program circuits extends from about 100 to about 5,000 cycles; to transmit music with improved fidelity a wider band than the above is desirable. A few circuits are at present available which extend the band down to 30 or 50 cycles and extend the higher range by 2,000 or 3,000 cycles. Program transmission circuits must be designed to handle wide ranges of volume. At present the volume range is limited to some 25 or 30 db, from about +8 vu down to about -22 vu. Obviously, since the dynamic range of a symphony orchestra is about 60 db, the wire-line circuit necessitates some compression of the dynamic range especially on long network circuits.

46. Standardization of Transmitting Levels. To obtain optimum conditions from the standpoint of noise and cross talk, it is desirable to transmit program material to loops at as high volumes as practicable. Telephone-company experience has demonstrated that in general +8 vu is about the highest volume of program material that can be tolerated in a local cable plant of the kind in which broadcasting loops are routed, from the standpoint of interference to other circuits. In view of these conditions, therefore, +8 vu (+14 vu output of amplifier followed by a 6-db pad) is the standard volume level for transmitting to loops in local telephone cables. The isolating pad is for the purpose of isolating the amplifier from the telephone company loops.

RADIO FACILITIES

47. A-f Equipment. The process of transferring programs from the main control room of the studios to the broadcast transmitting station is generally accompanied by a considerable reduction in the program signal level. Attenuation caused by the wire line upon which is added that caused by the line equalizer lowers the signal intensity as much as 25 db. A line equalizer consists of a specially designed network containing correctly proportioned values of L , R , and C . Irregularities in the wire-line frequency characteristics are smoothed out by the equalizer to produce a uniform frequency response of the wire line over as wide a range as practicable.

To increase the level of the incoming signals to a sufficient intensity to drive the

¹ Clark, A. B., *Wire Line Systems for National Broadcasting*. *Proc. IRE*, 17, 1998, November, 1929.

first tube of the speech amplifier of a broadcasting transmitter, a line amplifier is required. This amplifier is usually of a high-quality limiting type having sufficient gain to raise the audio program signal to a level of approximately +15 vu. At this level it enters the first speech-amplifier stage. The line equalizers, line amplifiers, variable attenuators, volume indicators, monitoring amplifiers, microphone for making local announcements, together with their switching equipment and jack panels, are normally mounted on shielded racks in a room called the *control room*. The shielding of the room itself sometimes consists of an outside-grounded copper screen containing within it a floating copper screen.

48. Limiting Amplifier. A special type of amplifier normally used in the speech-input layout at the broadcasting transmitter is of the compressing or limiting type. This amplifier automatically reduces the channel gain whenever the program peaks become excessively high. Thus it tends to prevent overmodulation. As a result, distortion due to transmitter overmodulation can be avoided while at the same time the average modulation can be raised with a corresponding audio power gain at the receiver. This is noticeable especially at low passages of program material where background noise may become objectionable.

By rectifying a small portion of the program signal output, a bias voltage control is provided on a program signal amplifier. This action does not just cut off the program peaks, but it reduces the gain and then allows it to again rise slowly to normal.

The signal voltage is amplified and then rectified in a diode with a result that a variable d-c bias voltage appears across a resistor in series with the bias voltage to the grids of the first stage of the amplifier. With an increasing signal, the bias becomes more negative and the output of the amplifier is reduced. This action does not occur, however, until the audio signal level applied to the control tube exceeds the fixed bias of this tube.

A potentiometer across the secondary of the input transformer is utilized as a variable-input control from which the corresponding input level at which the compression takes effect is varied. Owing to the high gain of the amplifier (60 db), the beginning of the compression may be as low as -40 vu. Provision is also supplied for adjustment of the output of the amplifier by means of a potentiometer in the input of the second amplifier stage. By means of this control the output level can be set anywhere within the range of -30 to +15 vu.

To compress sudden peaks of the program wave, the control circuit must function very quickly. The time constant of the circuits involved is such that the reduction in gain occurs in 0.001 sec. To prevent the gain from fluctuating at low audio or syllabic frequencies, there is a slow discharge or delay circuit provided to allow the compression bias voltage applied to the grids of the tubes in the first stage to leak off slowly and return the amplifier gain to normal in about 7 sec. This delay has been set by actual listening tests to prevent introduction of distortion or destroy speech inflections.

The amplifier has an output of +30 vu with 18 vu compression. The frequency response is flat within ± 1 db from 30 to 10,000 cps.

49. Program monitoring facilities are a very essential part of broadcast station equipment. In broadcasting technique, *program monitor audio* refers to a monitoring check on the audio signal input to the transmitter, whereas *program monitor radio* refers to a check on the demodulated signal secured by rectification of the carrier envelope as produced at the broadcast transmitter output. By switching from the *input signal* to that produced by rectification of the modulated transmitter carrier, the station personnel can determine by listening tests and measurements the relative amount of distortion produced in the broadcasting station equipment. For monitoring the outgoing program the personnel normally listens to the program monitor radio as produced by demodulation of the signal at the antenna system. This ensures that all portions of the audio and radio transmitting equipment, as well as the antenna system, are functioning. This is indicated by monitoring loud-speakers or oscillographs.

Facilities for program monitoring are provided in a room suitably constructed and acoustically treated to provide a favorable place for listening tests in the judgment of

quality. This may be either the transmitter room itself or an adjoining room called the control room where the speech input is normally located. The equipment for monitoring the audio signal consists of high-quality audio amplifiers, the gain of which can be regulated for proper signal volume; high-quality loud-speakers; and associated

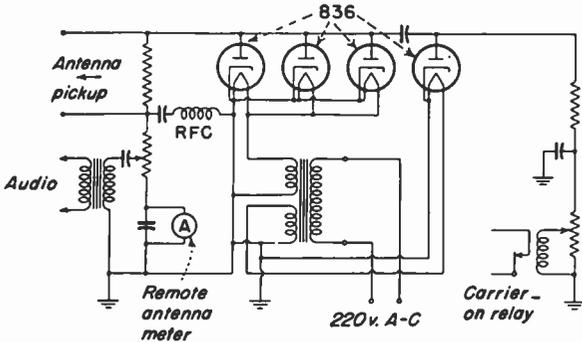


FIG. 52. Antenna monitoring rectifier circuit.

switching equipment. The frequency response of the entire system should be flat over a range of between 30 to 12,000 cps and higher. Additional to this equipment for program monitoring radio is a well-designed monitoring rectifier capable of demodulating the carrier signal as picked up at either the output tank circuit of the radio transmitter or at the antenna, preferably the latter.

Schematic diagrams of two types of antenna monitoring rectifiers, shown in Figs. 52 and 53, illustrate single-ended and push-pull types, respectively. These rectifiers are equipped with circuits enabling them to be used as the remote antenna current-meter rectifier, to close a carrier-on relay or time-outage clock relay as well as the monitoring signal for oscillograph or loud-speaker. In coupling such rectifiers as shown to an antenna circuit, precautions are usually taken to prevent the generation of even and odd r-f harmonics into the antenna circuit as produced by rectification. Under certain conditions such harmonic generation and radiation from the antenna system may create interference on the harmonic frequencies. For this reason, the push-pull type when inductively coupled to a high current point of the antenna system has considerable advantage over single-ended types, in that even harmonics are not as pronounced.

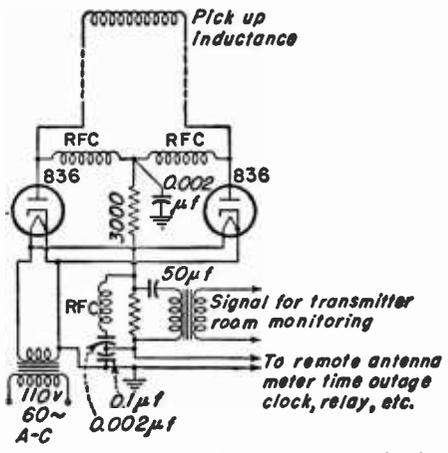


FIG. 53. Push-pull antenna monitoring rectifier.

For rectifying the envelope of a carrier wave to secure a signal for loud-speaker monitoring or for modulation measurements with an oscillograph, it is essential that the linearity characteristics of the monitoring rectifier between the impressed voltage and the plate current is substantially straight throughout the operating range. The unit must also have a uniform frequency-response characteristic to provide reproduction of the signal without frequency distortion.

Diode rectifier tubes are used extensively for monitoring radio telephone signals. As an individual element of the monitoring rectifier, the diode itself is not a linear device since the internal resistance of the diode decreases as the anode voltage is increased.

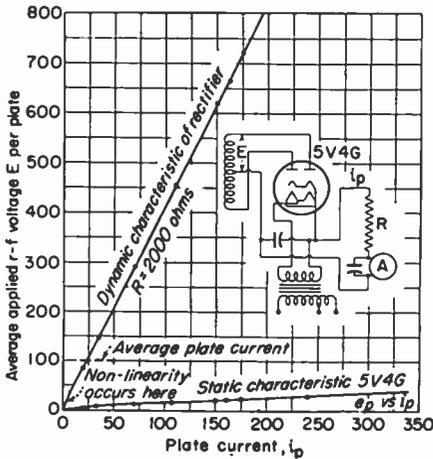


FIG. 54. Characteristics of monitoring rectifier. Total plate current of two plates is i_p .

from the dynamic characteristic by using a similar formula to that used in calculating percentage distortion of three-element tubes as audio amplifiers.

50. Frequency Monitor. This instrument is required at a broadcast station to measure the amount of deviation of the carrier from the assigned frequency. The FCC rules under Sec. 3.59 state that the operating frequency of each a.m. broadcast station shall be maintained within 20 cycles of its assigned value. Under Sec. 3.60 of the FCC rules, the frequency monitor is subject to FCC approval and must have a stability and accuracy within 5 ppm. The General Radio type 1181A frequency monitor illustrated in Fig. 55 is an approved type operating within the specified limits. It contains a frequency-standard oscillator utilizing a quartz crystal together with a tube (6SJ7) in a circuit having excellent frequency stability.

The oscillator drives a 6AC7 buffer amplifier with very light coupling between them. The output of the buffer is coupled to the grid of a 6SQ7 mixer where it is mixed with voltage from the broadcast transmitter. This voltage is picked up at some stage of the transmitter below the one modulated, is amplified by a 6AC7, and coupled capacitively to the mixer. The oscillator is adjusted to a frequency of 1,000 cycles off that of the transmitter carrier. Hence the output of the mixer is a difference frequency of 1,000 cycles plus or minus the transmitter deviation frequency. The frequency thus produced is amplified and passed through a 6SQ7 clipper-limiter tube to produce a square wave of constant peak-to-peak amplitude.

The square-wave voltage is amplified by a 6V6 tube circuit and passed to a voltmeter. This latter instrument is a full-wave differential diode voltmeter which indicates the potential difference between the mid-points of two series resonance circuits. In this manner equal potentials are indicated at exactly 1,000 cycles, and currents flow at other difference frequencies actuating the meter accordingly over the normal range of ± 30 cycles deviation which is the direct scale reading of the meter.

51. Modulation Monitor. Section 3.55 of the FCC rules requires that each broadcasting station shall have an approved modulation monitor in operation at the transmitter to measure the degree of modulation of the transmitter and for furnishing instant warning when the degree of modulation exceeds a selected specified value.

The selection of diode tubes having low internal voltage drop and the introduction of sufficient resistance in the plate circuit are required in the design of a monitoring rectifier of satisfactory linear characteristics. Linearity may be further improved by application of a constant positive bias in the plate circuit so that the diode draws steady plate current over the most nonlinear lower portions of the curves. In Fig. 54 these design features are illustrated for a 5V4G diode, which is a particularly good type for monitoring rectifier use due to its low internal voltage drop. Tubes of higher inverse peak voltage are often required for rectifiers of higher power handling characteristics and to withstand voltage surges (such as those caused by lightning) from an antenna circuit.

The percentage distortion of a rectifier may be approximately calculated

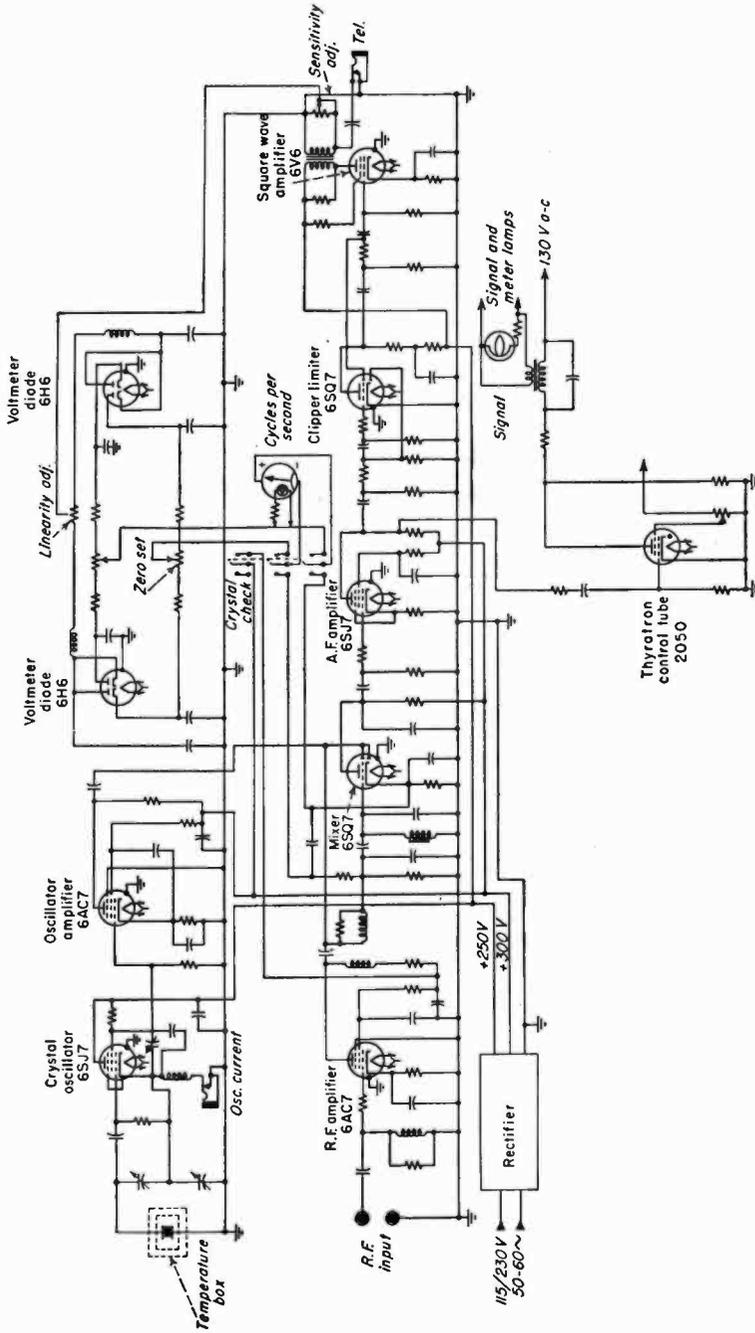


Fig. 55. Schematic of General Radio type 1181A frequency deviation monitor.

In Fig. 56 illustrating circuits of a modulation monitor, the amplitude-modulated r-f signal to be monitored impressed across diode 1-V is thus rectified. The resultant rectified pulsating d.c. flows in diode load resistor R_2 , R_3 . The average value of d.c. is indicated on the carrier meter. This is proportional to the average carrier voltage. The average component of voltage across this load resistor excites two indicating devices: (1) the meter M_2 calibrated to read modulation percentage and decibels directly and (2) a flasher and alarm circuit for providing a warning when the degree of modulation is exceeded.

The modulation indicating meter is excited in the following manner: The audio component secured from the first diode is rectified by the second 76 diode detector and charges C_5 . The voltage across C_5 is impressed across the grid circuit of the 76 vacuum-tube voltmeter stage which has the indicating meter M_2 in its cathode circuit. Circuit constants are made such in this instrument that the a-f peaks on the r-f carrier

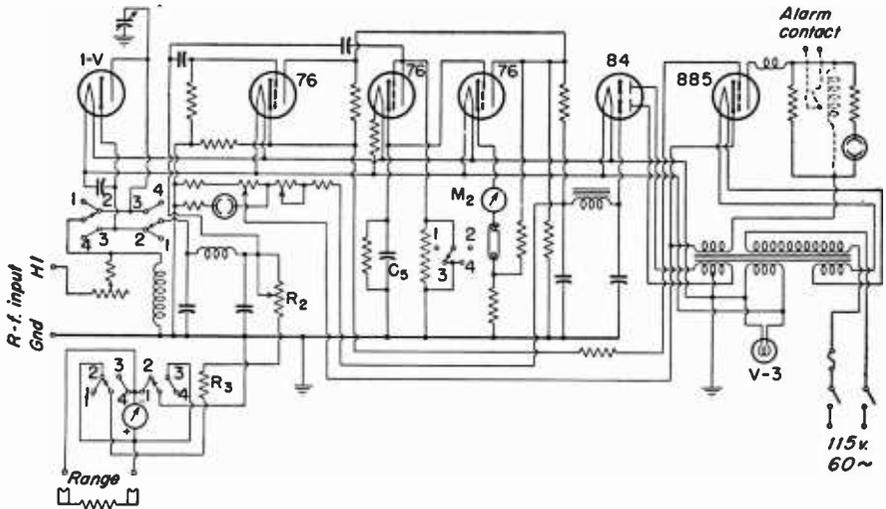


FIG. 56. RCA 66A modulation monitor.

are as indicated by meter M_2 . The neon flasher is operated by the relay tube, an 885, which is in turn driven from the first 76 tube under the same audio component from the 1-V tube that is used for operating the indicator meter system. If desired, the instrument may be used to operate an auxiliary alarm when the modulation peaks rise to an excessive value.

Modulation indicators are usually calibrated by means of a pure sine wave modulating signal applied to an accurate cathode-ray oscillograph and checked against the indicator. The frequency response must necessarily be flat over the audio range used to insure accuracy of measurement over the range.

According to Sec. 3.55 of the FCC rules, a license of a broadcast station will not be authorized to operate a transmitter unless it is capable of delivering satisfactorily the authorized power with a modulation of at least 85 per cent. When the transmitter is operated with 85 per cent modulation, not over 10 per cent combined a-f harmonics shall be generated. Under Sec. 3.46 (FCC rules) design recommendations call for the total a-f distortion from microphone terminals, including microphone amplifier, to antenna output should not exceed 5 per cent harmonics (voltage measurements) when modulating from 0 to 84 per cent and not over 7.5 per cent harmonics (voltage measurements of arithmetic sum) when modulating 85 to 95 per cent (distortion shall be

measured with modulating frequencies of 50, 100, 400, 1,000, 5,000, and 7,500 cycles up to the tenth harmonic or 16,000 cycles or any intermediate frequency that readings on these frequencies indicate is desirable).

The operating percentage of modulation of all stations is normally maintained as high as is possibly consistent with good quality transmission and good broadcast practice.

RADIOBROADCASTING TRANSMITTERS

Production of a broadcasting signal that will afford a means for conveying speech and music to the receiving set of a broadcast listener involves the generation of a constant r-f carrier upon which there are superimposed audio frequencies in the form of side bands, the intensities of which conform as nearly as possible with those contained in the sound produced in the studio. The production of such a signal may be accomplished by several methods¹ of modulation.

In American broadcasting technique the amplitude system of modulation is used exclusively in the present standard broadcasting band of 550 to 1,600 kc. The advantage of a.m. for transmission in this band lies in the production of a modulation envelope containing but a single pair of side bands, thus permitting station channel separation of 10 kc.

Compared with the a-m system, phase and frequency modulation produce an infinite number of side bands. It is evident that greater channel separation is needed and for this reason f-m stations have been assigned to the v-h-f part of the spectrum.

The primary requisites of a radio transmitter satisfactory to operate at a standard broadcasting station on an assigned frequency of between 550 and 1,600 kc under the present rules of the FCC for producing radiobroadcasting signals are as follows:

1. Satisfactory carrier frequency stability well within the allowable FCC tolerance of ± 20 cps maximum deviation.
2. Amplitude and frequency response characteristics providing low over-all signal distortion.
3. Suitable safety devices to avoid hazards to operating personnel and electrical circuits and equipment complying with the National Electric Code.
4. Minimum carrier noise level, approved electrical metering facilities; minimum r-f harmonic frequency power output; and freedom from parasitic frequency emissions.
5. Low operating costs requiring an over-all high-operating efficiency with respect to power input, low approved power-tube operating expenses and economical operating personnel requirements.
6. Durability, simplification of adjustment, and maintenance (requiring accessibility for repairs).
7. Reliability of service providing for continuous operation with a minimum of interruptions at rated carrier power output, modulated within legal limits.
8. Satisfactory dimensions for given power output providing for minimum installation and building costs.
9. Low initial transmitter and installation costs.
10. A pleasing appearance.

A recent trend is toward transmitters having high-level modulated and high-efficiency linear power amplifiers for the purpose of producing the desired high-quality broadcasting signal with a minimum of operating expense.

52. Typical Transmitting Equipment. In Fig. 57 is illustrated a simplified diagram of a radio broadcasting transmitter of recent design rated at 5-kw carrier power output such as that used at some standard broadcasting stations. It is commercially known as the RCA type BTA-10F.

The emitted carrier frequency of this radio transmitter is maintained well within a tolerance of ± 20 cycles by a crystal-controlled oscillator unit. This is accomplished through the use of V-cut quartz crystals having a temperature coefficient of about 1 part in 1,000,000 per degree centigrade. The mounting of the crystal is surrounded by

¹ Roder, Hans, Amplitude, Phase and Frequency Modulation, *Proc. IRE*, 19, 2145, December, 1931.

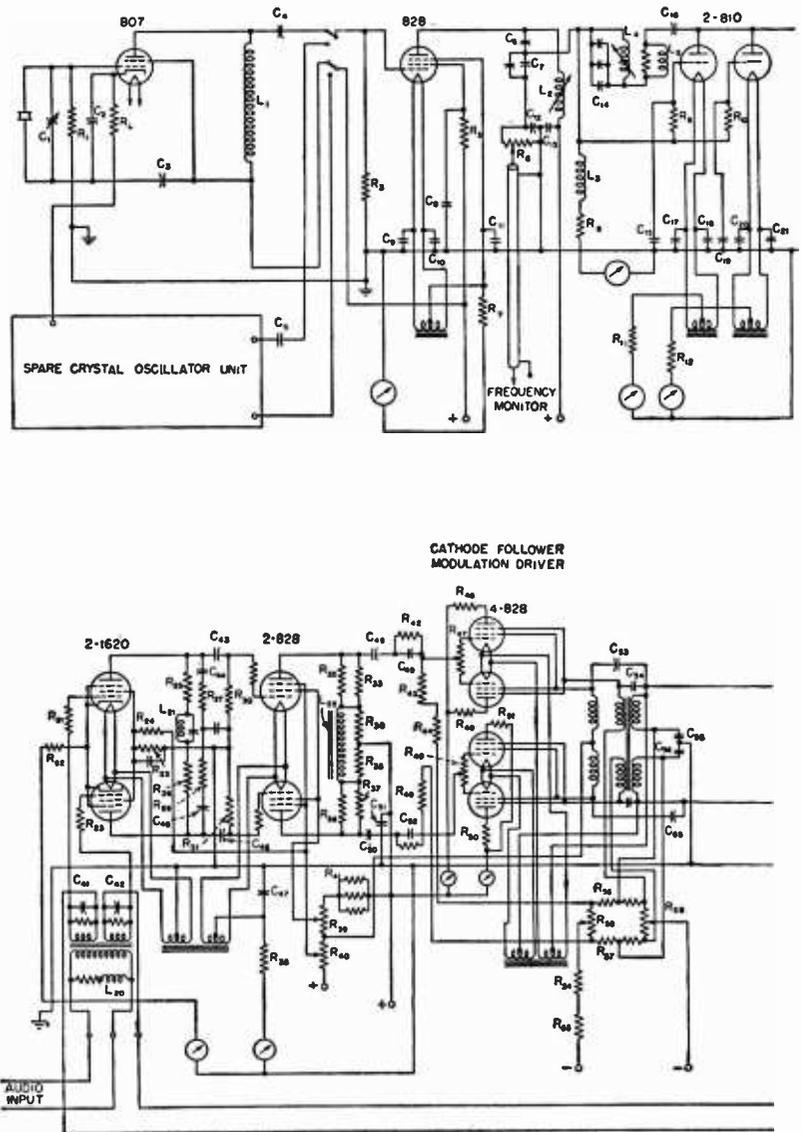
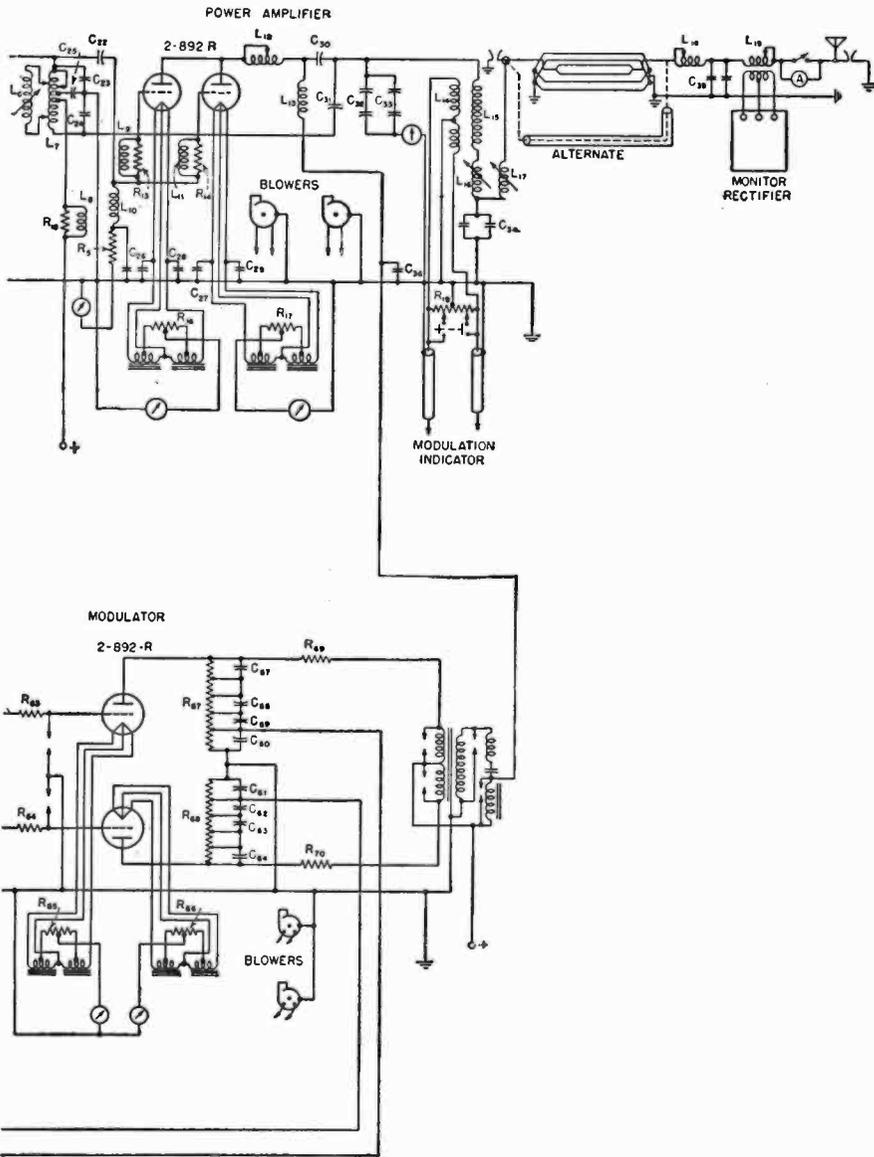


FIG. 57. Diagram of RCA type

a heater in close thermal contact with the bimetallic thermostat. The effects of changes in the ambient temperature are thus compensated, and the crystal is maintained at constant temperature. There is no tuned circuit associated with the crystal input circuit. Thus it is effectively "electron-coupled" to the output. A small trimmer capacitor is usually provided in shunt with the crystal to adjust it to "zero beat" or exactly to the desired carrier frequency.



BTA-10F broadcast transmitters.

Two crystal oscillator units are provided, one being a spare, which may be switched into use instantaneously. The output power of the crystal oscillator in use is amplified to the full 5-kw carrier output by a single 828 amplifier used to drive the parallel 810 intermediate power-amplifier stage. This drives the 892R power-amplifier stage. The modulated power amplifier is adjusted for plate-modulated class C r-f operation. The output of the power-amplifier stage is normally conveyed to the antenna by means

of a concentric or four-wire open transmission line through network circuits reducing r-f harmonic content to a very low value.

The transmitter utilizes high-level modulation; *i.e.*, the 892R stage is plate-modulated by a push-pull stage containing two 892R tubes. The modulator tubes are biased for class B audio operation for the purpose of securing high efficiency. The modulator is coupled by means of a modulation transformer to the plate supply voltage of the modulated amplifier. The modulator tubes are driven through an input transformer by four 828 tubes operated push-pull parallel in a cathode-follower circuit. The 1620 and 828 stages are operated as class A audio amplifiers.

Elaborate precautions have been taken in the design of the audio stages to control the phase rotation with respect to the frequency characteristic in those circuits to which degenerative feedback has been applied. All circuit elements, especially the audio transformers, must have a minimum of phase shift over the a-f range to realize advantages from the application of degenerative feedback. As illustrated in Fig. 57, a potentiometer across the primary of the modulation transformer provides a signal voltage that is introduced out of phase into the input of the audio system. Hum or noise generated in the r-f power amplifier appears across the modulation transformer and is thus also introduced out of phase to the speech amplifier input. Therefore, with degeneration, the over-all carrier noise level is very low. Measurements indicate this to be 60 db below the signal level of 100 per cent modulation. The amplitude distortion is maintained by this system well below 3 per cent rms over the a-f range of between 30 to 10,000 cps, and the over-all frequency response of this transmitter is substantially flat within $1\frac{1}{2}$ db over this audio range.

Features of this transmitter which merit consideration are its simplicity brought about through the use of a-c filament supply for all tubes, thus eliminating filament motor-generator sets. This points toward a considerable saving in power and vacuum-tube operating costs as well as on transmitter space requirements and initial installation costs. Reduction of carrier noise level of this transmitter to an extremely low level is accomplished through the use of indirectly heated cathodes of tubes in the low-level stages and the use of degenerative feedback. The transmitter requires no water-cooling system since all power tubes are air-cooled.

A very small portion of the modulated r-f power produced by the power amplifier is introduced with proper phase rotation into the first audio stage of the transmitter to reduce carrier hum and noise. Design features require minimum phase shift in all circuits involved to permit satisfactory operation of this system of reducing carrier noise and envelope distortion.

A transmitter may be supplied with a phantom antenna for use during transmitter warm-up and test periods. Switches are provided for transferring r-f carrier power from the output stage of the transmitter to either the radiating antenna or the phantom. The latter is designed to act as an effective resistance-load equivalent to the characteristic impedance of the transmission line. It must necessarily be capable of dissipating 15 kw of r-f energy in a 10-kw transmitter when modulated 100 per cent with a sustained audio signal having sinusoidal wave shape.

Figure 58 illustrates a simplified circuit of a Western Electric type 443A-1 transmitter capable of delivering 1 kw of 100 per cent modulated carrier output power to an antenna. The modulation system consists of grid-bias modulation applied to the r-f output stage utilizing four 357A tubes in a Doherty high-efficiency circuit. In view of the small amount of audio power requirements from the modulator of this system, the audio and modulator stages are quite simple and of low power. Stabilized feedback is used between the power amplifier output stage and the first audio stage of the speech amplifier to reduce over-all distortion and carrier noise. The rms a-f harmonic distortion over the range from 50 to 3,000 cycles is less than 3 per cent at 100 per cent modulation and over the range to 7,500 cycles it is less than 5 per cent. The rms noise level is normally of the order of 70 db below a signal produced by a 100 per cent modulated carrier. The frequency response is flat within 1.5 db from 30 to 10,000 cycles.

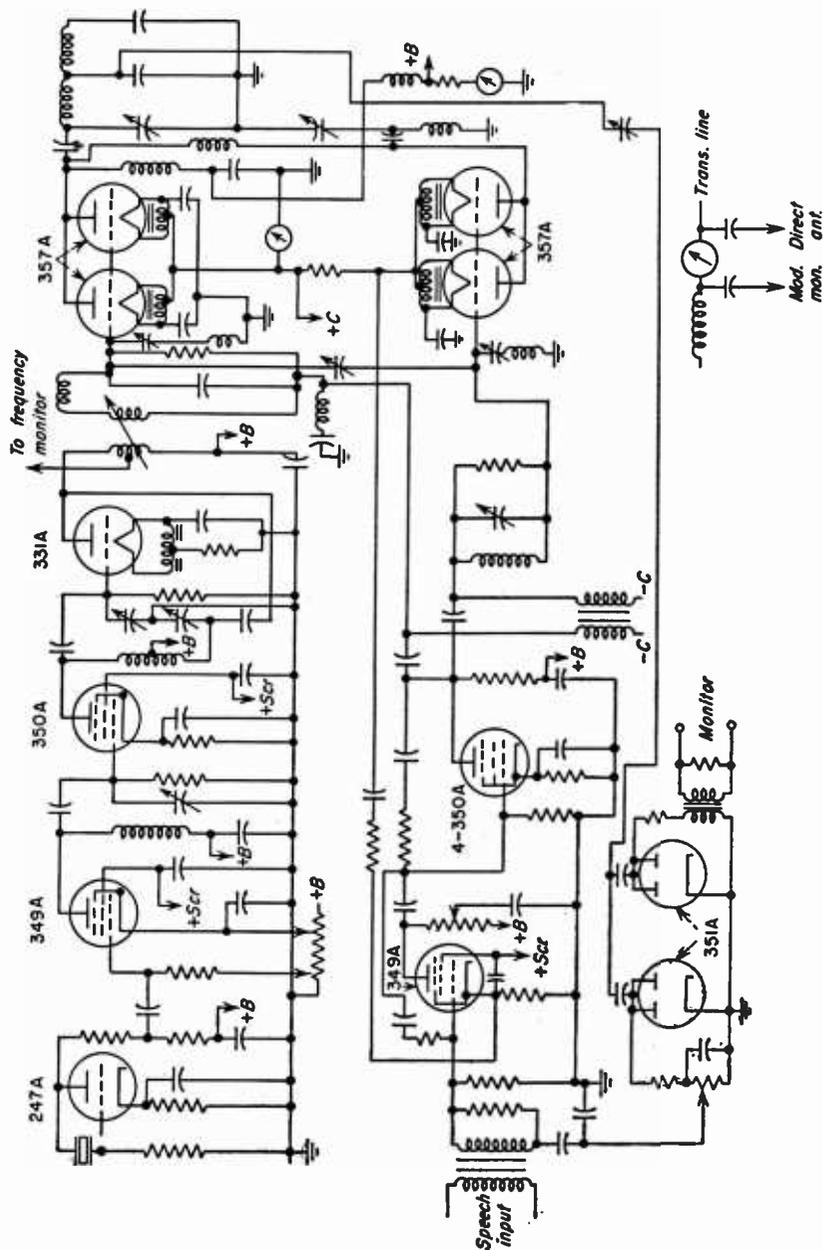


FIG. 58. Simplified circuit of Western Electric 443A-1 transmitter.

The over-all power consumption is 4.3 kw at 85 per cent power factor for an unmodulated carrier output of 1,000 watts, and it is 5.15 kw for a single-frequency 100 per cent modulated carrier output.

53. International Broadcasting. Transmitters for this service are operated at high frequencies and for this reason are considerably different in design from transmitters operated in the 550- to 1,600-kc band. They are used with directive antennas having a power gain of 10 or more, and have carrier powers up to 200 kw 100 per cent modulated.

R-F CIRCUITS

54. R-f Amplifier Neutralization. One of the essential adjustments in an r-f amplifier circuit to obtain stability and prevent self-oscillation is accomplished through neutralizing the electrostatic capacitance of the grid-to-plate electrodes in the triode power tubes.

For the purpose of neutralizing an amplifier stage such as the final r-f power amplifier shown in Fig. 57, first remove plate voltage from it and apply normal r-f excitation to the grid circuit. Tune the grid circuit to resonance in the usual manner. Next connect a low-power (5- to 10-watt) high-resistance lamp across one or two turns of the plate tank inductance. The leads to the lamp should be very short and provided with clips for convenience. Next tune the plate tank circuit to resonance with the grid exciting voltage frequency as indicated by maximum brilliancy of the lamp. It is to be noted that the circulating current in the plate tank circuit which lights this lamp includes the coupling effect of the grid-plate capacitance of the tube.

The neutralizing voltage of opposite polarity is obtained by connecting to the opposite end of the grid or plate tank circuits, as the case may be. The magnitude of the voltage used to neutralize the grid-plate capacitance current is regulated now by adjusting a neutralizing capacitor. As the neutralizing capacitor is varied, the lamp will change brilliancy, and, when correct balance is obtained, the lamp will be at practically zero brilliancy. As neutralizing capacitance is changed, some slight corrections in plate-tank tuning and in grid tuning may be necessary, owing to interactions of these two circuits. Always tune to resonance by maximum lamp brilliancy and neutralize for minimum brilliancy.

When best results are obtained by the lamp method, remove it from the plate coil, and, if more accurate adjustment is required, a low range r-f ammeter should be inserted in series with the tank circuit. By using a meter, maximum accuracy is obtained by tuning the circuit to obtain absolute minimum current.

Since the effect of coupling between successive stages greatly affects the neutralizing, the adjustment should be made with all circuit conditions and couplings as nearly final as possible.

An entirely different neutralizing system consists of an effective inductance shunting the interelectrode grid-to-plate capacitance of the power amplifier tubes. Suitable d-c blocking capacitors are provided to prevent the plate voltage from reaching the grid through this neutralizing inductance.

Neutralizing adjustments with this shunt inductance may be accomplished with a high resistance lamp or thermomilliammeter attached to the output tank circuit in much the same manner as was described for capacitor neutralization except that neutralization is accomplished by adjustment of the shunt inductance. This system has great advantages over the neutralizing capacitor method especially where it is desirable to keep circuit tank capacitance and the corresponding kva/kw ratio to a low value. This is the case where stabilized degenerative feedback is applied through an amplifier stage wherein a minimum phase rotation with frequency is required.

55. Class B Linear R-f Amplifiers. The operation of a push-pull class B r-f amplifier may be understood by a study of Fig. 59. Here it is shown that plate current drawn by the tubes is very closely a linear function of the grid-voltage swing. The

associated output-circuit loading is adjusted so as to realize from the tube a maximum conversion efficiency. Some curves showing how plate-current efficiency varies with effective load impedance are shown in Fig. 60. The crest position on these curves depends upon the tube characteristics and the power factor of the circuit into which it operates. These curves were taken at a standard broadcast frequency by varying the load upon the output circuit of a linear amplifier stage and measuring the efficiency of the stage at various d-c plate voltages.

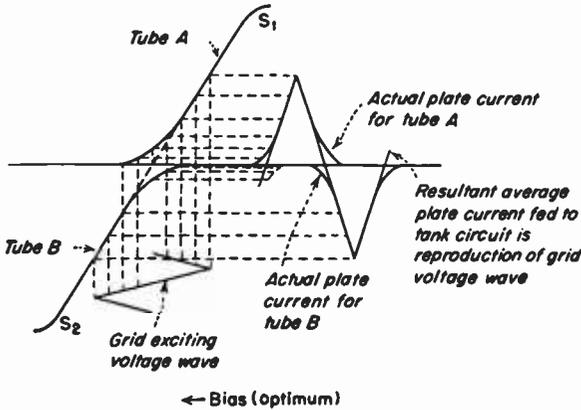


FIG. 59. Theoretical curves showing push-pull class B amplifier operation.

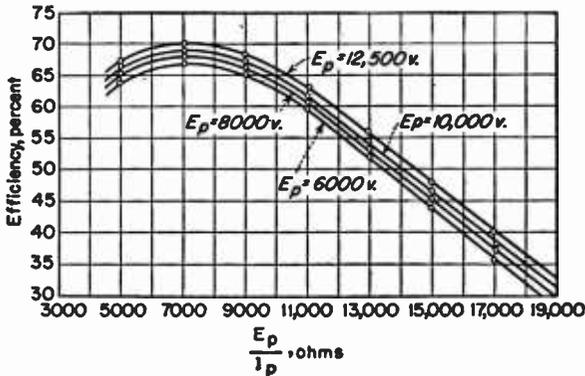


FIG. 60. Load characteristics of push-pull 892 power amplifier tubes.

Under conditions where the conversion efficiency is a linear function of the grid swing, the power output is necessarily proportional to the square of the grid swing. Hence the peak power output at 100 per cent modulation is four times that at which the modulation is zero. The steady power output under conditions of sustained 100 per cent modulation is 1.5 times the output of zero modulation. Therefore, in considering power-tube requirements for a class B linear-amplifier stage, provision must be made with respect to filament emission and plate dissipation so that the tubes are capable of supplying peak power outputs of four times that of the nominal carrier-power output rating of the transmitter. This assumes that the modulation capability of the transmitter is 100 per cent.

In adjusting a push-pull linear amplifier, both sides of the circuit must necessarily have very nearly identical operating conditions with respect to grid swing and circuit adjustment, so that equal plate currents are measured on the individual tubes identified as *A* and *B* in Fig. 59. The grid-bias adjustment necessarily depends directly upon the plate voltage used since the position of the characteristic curve is moved with each corresponding change in plate voltage.

As illustrated in Fig. 59, with a simple triangular wave form, the method of determining optimum grid bias depends upon the point where an extension of the straight portion of the curve intersects the horizontal axis. Both dynamic curves of tubes *A* and *B* have their straight portions in direct alignment. Distortion due to the lower bend in each characteristic curve is averaged out together with the *kva/kw* inertia effect in the output-tank circuit. On the other hand, it is illustrated in the curves that, for maximum modulation peaks with output increasing as the excitation voltage is increased, there is a limit to the output as represented by the upper bends, points *S*₁ and *S*₂, on the curves where the tube saturation points begin.

Linearity is therefore dependent upon grid bias, grid-exciting voltage, and output-tank loading. The procedure for setting taps for correct output-tank loading consists of first saturating the grids of the amplifier tubes with sufficient r-f grid driving power. Then with one-half normal class B r-f operating plate voltage applied, the amplifier is loaded until it delivers rated carrier power normally to the antenna, the plate efficiency being usually between 65 to 70 per cent. Then the grid-exciting voltage is reduced (usually by means of grid-loading resistors) until the amplifier stage with full plate voltage applied delivers the same rated carrier output with a corresponding plate efficiency of very nearly 35 per cent. This is the plate efficiency for a class B r-f amplifier as specified by Sec. 3.52 of the FCC rules in the determination of carrier output power by the indirect method. Under this same section the plate efficiency for plate-modulated class C r-f operation of the last radio stage as measured by the indirect method is 70 per cent for transmitters having carrier power output up to 1 kw and 80 per cent for 5 kw and over.

56. High-efficiency linear-power amplifiers are a result of work to reduce the expense for operating power of broadcasting transmitters usually of 1-kw carrier output and above. The limitations of the class B r-f linear amplifier as previously discussed illustrate that for satisfactory operation of this system the plate power efficiency ranges from 30 to 35 per cent for the stage. Considering the driver and modulator stages and the transmitter auxiliaries, with this system the over-all efficiency from power mains to carrier power output may range from 20 to 25 per cent. The high-efficiency amplifier circuit¹ provides a plate operating efficiency of as high as from 60 to 65 per cent to be realized from a linear-power amplifier.

The amplifier circuit (Fig. 61) has been divided in block form into individual units. Voltages at points in the circuit are as indicated by symbols *E*_z, *E*₁, *E*₂, *E*₃, and *E*₀.

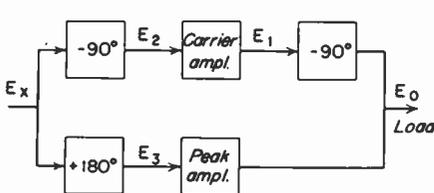


FIG. 61. Block diagram of high-efficiency power amplifier.

The exciting voltage *E*_z passes into two branches. One leads into a negative 90-deg phase-shifting circuit, thus transforming it to the proper amplitude for grid excitation of the carrier amplifier tube. This grid voltage *E*₂ is amplified by the carrier tube, the a-c components of plate voltage becoming *E*₁ (180 deg out of phase with *E*_z). The output voltage *E*₁, in passing through the impedance-inverting network shown has its

phase retarded an additional 90 deg at the output of the network. Therefore, in turning through 360 deg in this path, the resultant *E*₀ is in phase with the exciting voltage *E*_z. In the lower branch of the circuit the 180-deg phase reversal of *E*_z in passing

¹ Doherty, W. H., A New High Efficiency Linear Amplifier for Modulated Waves, *Proc. IRE*, September, 1936.

through the grid network and the phase reversal produced in passing through the *peak amplifier* tube results in a correct phase of E_o at the load. The phase shifts may be further clarified by the vector diagram of Fig. 62, where the output voltage produced by both the carrier and peak tube are illustrated as acting in phase to produce E_o at the load.

Figure 63 illustrates the theoretical individual and combined operation of the carrier and peak tube branches of the power amplifier unit as produced by a modulated r-f exciting voltage (assuming sinusoidal variation of r-f voltage with modulation). For the carrier amplifier tube the r-f plate voltage rises very nearly linearly over the region O to A , flattening off at this point due to saturation; beyond this point any increase in grid-exciting voltage for this tube produces practically no further increase in plate voltage. The carrier amplifier tube plate current on the other hand rises quite linearly from O to B . Thus from O to A the operation is quite similar to that of a class B r-f linear amplifier operating into a load impedance of constant value; whereas from A to B there is a progressive reduction in its plate impedance under influence of positive delivery of power from the peak amplifier tube on upward modulation swing, as observed through the impedance-inverting network and the plate-current rises. The plate voltage of the peak amplifier tube rises linearly from O to B , where the curve flattens off because of saturation. This tube is biased to a point where little positive

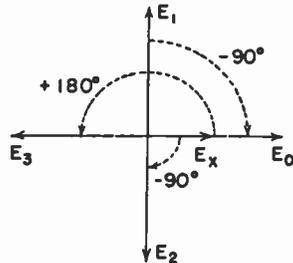


FIG. 62. Phase relations in high-efficiency amplifier.

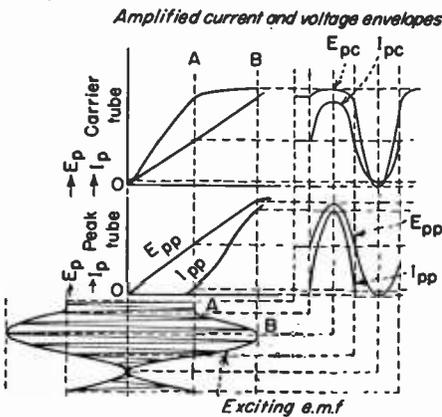


FIG. 63. Operating characteristics of high-efficiency amplifier.

power is delivered for grid-exciting voltages below carrier amplitude A . However, owing to coupling to the carrier amplifier tube output circuit through the impedance-inverting network, a voltage exists in its plate circuit during this idle stage for the tube. Therefore, a linear variation in plate voltage for the carrier amplifier tube between O and A causes a corresponding linear voltage variation between O and A in the plate circuit of the peak amplifier tube because it is in parallel with the load. Owing to grid-biasing conditions with respect to the excitation voltage for the peak amplifier tube, appreciable plate-current flow begins when exciting voltage assumes an amplitude greater than that necessary for an unmodulated carrier condition. Over the region A

to B , plate current rises very nearly linearly to the limit at B . It is evident that at the crest of the modulation cycle corresponding to B both the carrier and peak amplifier branches are delivering equal power outputs in phase to the load. Adjustments required for satisfactory operation consist of correct neutralization of the tube capacitance, correct grid biasing of carrier and peak amplifier tubes, adjustment of grid load resistors of both amplifier tubes and their grid and output tank circuits to resonance, as well as obtaining correct phase-inverting characteristics from the circuits involved. For correct loading of the amplifier, the r-f transmission line should be properly terminated to permit operation of the amplifier into a resistive load. It will be noted that for the purpose of securing the impedance-inverting characteristics required, a 90-deg phase shift is also secured. All other phase-shift networks are

utilized to compensate for this undesired phase shift. Compensation for phase shift must be effective over all useful side-band frequencies and also at the carrier frequency. The 90-deg phase-shifting circuit in the grid of the carrier amplifier tube and the 180-deg phase-shifting circuit in the grid of the peak amplifier tube are utilized for compensation purposes only.

57. Stabilized degenerative feedback as applied to radiobroadcasting transmitters reduces the audio-harmonic distortion and noise created within the transmitter equipment, thus providing high-fidelity performance. Reduction of carrier-noise level may be carried to as low as 65 db below 100 per cent modulation signal by utilizing degenerative feedback, even with a-c applied to the filament of all tubes. The FCC, Sec. 3.46, recommends that the carrier hum and extraneous noise (exclusive of microphone and studio noises) level (unweighted rms¹) be at least 50 db below 100 per cent modulation for the frequency band of 150 to 5,000 cycles and at least 40 db down outside this range. Harmonic distortion may be reduced to well below the FCC requirements, and in some cases the measured value of rms a-f harmonic distortion in the range 50 to 5,000 cycles is less than 2 per cent at 85 per cent modulation and less than 3 per cent at 100 per cent modulation even with a high-efficiency power amplifier unit as a part of the system.

The application of degenerative feedback to the transmitter (Fig. 57) follows the same principles as applied to audio amplifiers. Application of feedback to radio transmitters is, in general, more complex than when applied to amplifiers. Theory shows that, if a part of the output of an amplifier or radio transmitter is fed back to the input and combined with the input signal in reverse phase, the effective gain is reduced. However, if the signal fed back contains noise and distortion components not present in the input signal, these components will be amplified by the full gain of the amplifier and, in traveling through the system to the point where they were picked up, will tend to neutralize the distortion and noise in the system provided that the fed-back signal is exactly 180 deg out of phase with the input signal and the phase shift through the system is small over the range of the distortion frequencies. Under such conditions the distortion will be reduced in amplitude by the amount of gain reduction.

In operating a transmitter with feedback the over-all gain of the audio system is reduced by the amount of feedback used. For example, if 30 db of feedback is employed and the feedback voltage is removed suddenly by some fault, the program input will be 30 db too high, and readjustment of the program input level must be made instantaneously to prevent overloading. In the transmitter of Fig. 58 the feedback voltage is secured by rectifying a small portion of the power output of the power amplifier unit by means of a feedback rectifier designed for minimum phase shift. This voltage is introduced into the first speech amplifier audio stage together with the audio input signal.

With the application of degenerative feedback to cascade r-f amplifiers it becomes extremely difficult to maintain the phase of the rectified signal picked up at the output of the transmitter sufficiently close to the 180-deg rotation required throughout the entire a-f range. Unless all networks in the entire cascade system are correctly designed, the kva/kw ratio of all r-f tank circuits are kept to a very low value, and stray capacitances are minimized, there is an accumulative phase shift through the feedback loop wherein the degenerative system is active.

Under conditions where the voltage fed back to the audio input of the transmitter after passing through the feedback loop is other than 180 deg out of phase with the input signal, less noise and distortion cancellation result. This is especially true under conditions where the phase shift of the feedback loop becomes less than 90 deg or more than 270 deg. At frequencies where the phase shift approaches zero and 360 deg from that of the input signal, stabilizing circuits are necessary to prevent violent oscillation of the entire transmitter at these frequencies, provided, of course, that the amplification around the loop is at least unity. These frequencies are sometimes referred to as those at which the phase "turns over." For the h-f turnover point, say around 25 kc,

¹ Root sum square.

an adjustable stabilizing filter may be utilized in one of the low-power speech amplifier audio stages. This prevents oscillation or singing of the transmitter at the particular high a.f. where the condition exists and for this reason is called the "anti-sing" circuit. In addition there may be required a l-f (below 100 cycles) stabilizing circuit in one of the low-power audio stages to prevent oscillation at the l.f. at which another unstable condition exists.

By correct proportioning of all constants of the a-f and r-f stages and associated networks throughout the entire section of the transmitter containing the feedback loop and by application of the stabilization circuits as mentioned together with careful transmitter adjustments, an effective amount of feedback can be normally secured for cancellation of noise and distortion.

MODULATION EQUIPMENT

58. Speech Amplifier. An audio-amplifier unit employing power tubes is usually necessary as the preliminary part of the audio system of a transmitter to raise the audio-signal intensity to a sufficient amount to swing the grids of the modulator tubes. Resistance coupling is frequently used in speech-amplifier circuits. In Figs. 57 and 58 are shown simplified circuit connections of typical transmitter speech amplifiers.

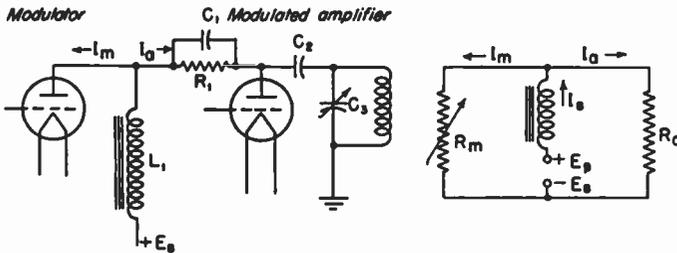


FIG. 64. Heising constant-current modulator and equivalent.

Amplitude modulation provides a means for reproducing a signal containing a distortion not exceeding a few per cent with the carrier fully modulated. In broadcasting transmitters it can be effected by either plate or grid modulation. When grid modulation is applied to a power amplifier tube, either by bias-voltage or r-f grid-voltage change, the efficiency of the power amplifier is rather low, ranging from 30 to 35 per cent. A plate modulated radio stage operating as a class C amplifier has a comparatively high efficiency ranging from 70 to 80 per cent. This advantage of higher efficiency, however, is offset by the low efficiency of the plate modulator unless a class B audio amplifier is used for modulating. Therefore there is not much difference in the two systems, in so far as efficiency is concerned, with respect to power and vacuum-tube costs except under conditions where modulating power for a class C r-f output stage is supplied from a modulator of rather high efficiency.

When the power-amplifier stage of the transmitter is plate-modulated, the setup is called a *high-level system* of modulation; whereas a transmitter modulated in a low-power stage of the transmitter and followed by a class B r-f power amplifier is termed the *low-level system* of modulation.

59. Modulators and Modulated Amplifiers. In Fig. 64 is shown a constant-current system of modulation due to Heising.¹ The modulator and modulated amplifier are connected in parallel with a constant-current source of supply. This is connected to the common plate lead through a large inductance L_1 called the *modulation choke*.

The dynamic modulating characteristics can be determined with a fair degree of accuracy from the static characteristics of the modulator tubes in a method illustrated in Fig. 65. The modulated amplifier is assumed to be a pure resistance load in parallel

¹ Heising, R. A., Modulation in Radio Telephony, *Proc. IRE*, 9, 365, August, 1921.

with the plate resistance of the modulator tubes and both assumed to be supplied with power through a modulation choke of infinite impedance. The sum of the instantaneous currents in the amplifier and modulator in this case is a constant. An approximation is made of the number of modulator tubes required to modulate a given r-f amplifier. The plate-current ordinate for a single tube must be multiplied by the

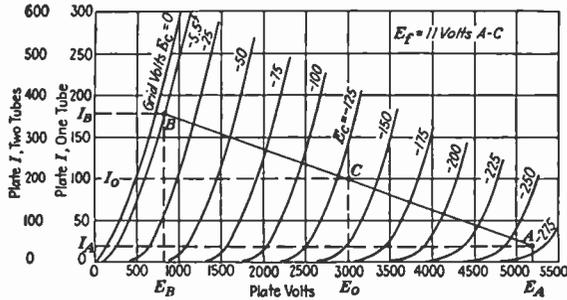


FIG. 65. Method of determining modulator characteristics.

number of modulator tubes before the load line *BA* can be plotted, the slope in amperes per volt which depends upon the load resistance produced by the amplifier. Line *BA* was chosen for two modulator tubes operating at 3,000 volts plate into an amplifier of 2,000 volts and 150 ma or an effective resistance of 13,333 ohms. The mean modulator plate current *I*₀ is chosen from allowable

plate dissipation and load line *BA* drawn in about operating point *C*. The modulator grid voltage swings from $-\frac{1}{2}E_f$ (filament voltage) to equal grid voltage on the other side of the operating point. By taking readings of plate current and voltage from end points of the load line, the following information becomes available:

$$\text{Modulation factor} = \frac{E_A - E_B}{2E_0}$$

where *E*_A = max plate-voltage swing

*E*_B = min plate-voltage swing

*E*₀ = d-c plate voltage at operating point *C*

$$\text{Per cent second harmonic distortion} = \frac{\frac{1}{2}(I_A + I_B) - I_0}{(I_B - I_A)} \times 100$$

where *I*_A = max plate-current swing

*I*_B = min plate-current swing

*I*₀ = plate current at operating point *C*

$$\text{Power output in watts} = \frac{1}{8}(E_A - E_B)(I_B - I_A)$$

60. Design for High Audio Fidelity. In the design of the modulated amplifier circuit of the above system certain elements of the circuit must be properly proportioned to afford a uniform frequency characteristic. The capacitance of *C*₁ (Fig. 64) should be large enough so that its impedance at the lowest frequency to be transmitted is less than one-third of *R*₁, or the plate-dropping resistor.

Capacitor *C*₂ provides an r-f path from plate to filament of the amplifier tube and at the same time breaks the d-c path. It must also break the path for higher frequency a-f current and permit it to flow through the amplifier tube. It should, therefore, be no larger than necessary to conduct the r-f plate current without producing excessive phase shift in the plate current under conditions where *C*₂ is less than 2*C*₃.

Sufficient impedance of the modulation choke over the a-f range is another important factor in circuit design. Its impedance at the lowest a.f. should be at least two times the effective resistance load produced by the r-f amplifier tube. The choke should be free from inherent self-capacitance defects over the frequency range to maintain a sufficiently uniform high impedance at the higher frequencies.

High-quality signal reproduction requires that amplitude distortion should be kept at a minimum. A common cause of amplitude distortion is due to underexcitation of the grid of a modulated amplifier tube when plate modulation is applied. This results in insufficient driving voltage during periods of high plate-voltage swing and conse-

quently peak-output limiting. Trouble from this cause shows up quite clearly upon an amplitude curve or upon an oscillograph in the form of chopped-off positive peaks. In Fig. 66 are shown amplitude curves taken on the modulated carrier of a stage the grid of which was excited to saturation as shown in *A* and underexcited in *B*. It is a custom to have available a surplus of driving power for a modulated amplifier to prevent any possible occurrence of amplitude distortion.

The constant-current or Heising system of plate modulation is often designated as a class A system, since the modulator tube performs under conditions similar to those encountered in a class A amplifier.

A properly designed class B system permits a much higher plate efficiency to be secured from a given set of tubes and correspondingly a much greater output from them than with a class A system. This efficiency has been made to reach as high as 66.6 per cent with a small percentage of audio harmonic distortion.

Inasmuch as it is often necessary to drive the grids of class B audio amplifiers into their positive grid-current region to obtain maximum power output, it is important that the driver-amplifier stage for the modulator stage should have a good output-voltage regulation. This calls for driver tubes having a sufficient output capacity to deliver an undistorted voltage to the grids of the class B stage, even though there is a nonuniform increase of load on the driver stage caused by the class B tubes as they are driven through the positive grid-current region of their dynamic operating characteristics.

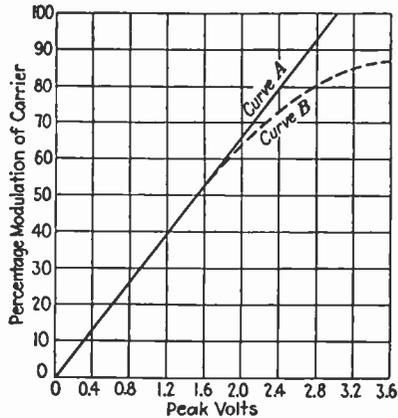


FIG. 66. Amplitude curves taken on a modulated amplifier. Curve *A* taken on stage with sufficient driving power applied to saturate grid. This shows negligible amplitude distortion. Curve *B* taken on stage with insufficient grid excitation to cover positive peaks. Amplitude distortion becomes noticeable at 60 per cent modulation and increases with higher levels.

F-M SYSTEMS

The method of program signal transmission by means of f.m. utilizes a frequency variation or deviation at the audio rate, the deviation frequency being a small percentage of the unmodulated carrier frequency.

Assume the existence of an f-m transmitter operating on 90 Mc and that a maximum deviation of ± 75 kc is desired. Then a sustained sine wave of, say, 1,000 cps may be applied to the modulator audio input, the amplitude of the audio signal adjusted to provide ± 75 kc deviation. This would result in utilizing the full modulation capabilities of the transmitter. With a complex wave program input, the frequency deviation at any instant corresponds to the amplitude of the complex wave at that instant.

Channels for f-m transmissions have been assigned 200 kc apart, which has been found to be a sufficient carrier separation to allow a frequency deviation of as much as ± 75 kc. The width of the band required¹ in the frequency spectrum is at least twice the value of the highest modulating frequency or twice the frequency deviation, whichever is greater. Important side-band components may occur outside these limits, however.

61. **Methods.** There are diverse methods of producing f.m. on an r-f carrier. Two rather different systems have been classified as (1) direct f.m. and (2) indirect f.m., accomplished primarily by phase modulation.

¹ Carson, J. R.. Notes on the Theory of Modulation, *Proc. IRE*, February, 1922; van der Pol, B., Frequency Modulation, *Proc. IRE*, July, 1930.

Direct f.m. is produced by frequency-modulating directly the master oscillator stage, which has a normal unmodulated carrier frequency of either the transmitter output frequency or a convenient subharmonic thereof. In papers¹ giving a mathematical treatment of f.m., it has been illustrated that, if the tank circuit constants of the oscillator stage are varied in accordance with the audio-input frequencies, there will be produced a resultant f-m output signal. With some device operating as a condenser microphone varying the capacitance of the tank circuit of the master oscillator, there

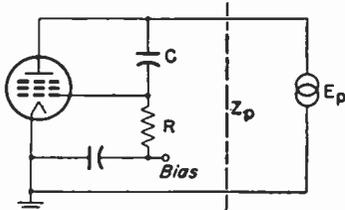


FIG. 67. Reactance-tube modulator.

may be produced an r-f carrier frequency modulated to conform with the sound undulations vibrating the microphone diaphragm. This illustrates f.m. by the direct method.

A modified form² of the direct system of frequency-modulating a transmitter is accomplished through the use of a tube (Fig. 67), employed as a variable reactance. Here a variable reactance is caused to exist between the cathode and anode of the reactance tube by grid-bias variation at an audio rate. By supplying the grid of the reactance tube with r-f voltage previously passed through a phase-shifting circuit of suitable resistance and capacitance, the grid-excitation voltage is caused to be in phase quadrature with the plate voltage. Then the a-c portion of the reactance-tube plate current i_p will be very nearly

$$i_p = g_m e_o$$

However, since

$$e_o = jK e_p$$

then, under influence of the phase-shifting network,

$$i_p = jK g_m e_p$$

and

$$Z_p = \frac{e_p}{i_p} = -\frac{j}{K g_m}$$

From which the equivalent capacitance produced by the tube

$$C_e = \frac{K g_m}{2\pi f}$$

The reactance tube may be caused to appear as an equivalent variable capacitance across the oscillator tank circuit.

62. Carrier Stabilization. For satisfactory transmission of the f-m signal it is essential that, in addition to producing the modulated wave, there must be present a satisfactory carrier-wave stability. For this reason it is necessary to add a stabilizing circuit by means of which the average frequency of the carrier wave is compared to that of a precision crystal oscillator. The resulting difference frequency when applied to a special motor coupled to a tuning compensator provides carrier-frequency correction to the assigned channel frequency.

In Fig. 68 are shown the basic circuits for monitoring the transmitted carrier frequency within the required frequency tolerance for an f-m broadcast transmitter. The system consists of two balanced modulators arranged to secure in the output of each a beat frequency equivalent to the difference between a crystal-frequency standard and that of the master oscillator. The output of the crystal oscillator is divided to feed equal amounts of energy to the two balanced modulators. Two phase-shifting networks provide a 90-deg phase displacement between the input signals applied to the two balanced modulators.

¹ Roder, H., Amplitude, Phase and Frequency Modulation, *Proc. IRE*, December, 1931. Crosby, M. G., Frequency Modulation Propagation Characteristics, *Proc. IRE*, 24 (6), June, 1936.

² Crosby, M. G., Frequency Modulation Noise Characteristics, *Proc. IRE*, April, 1937.

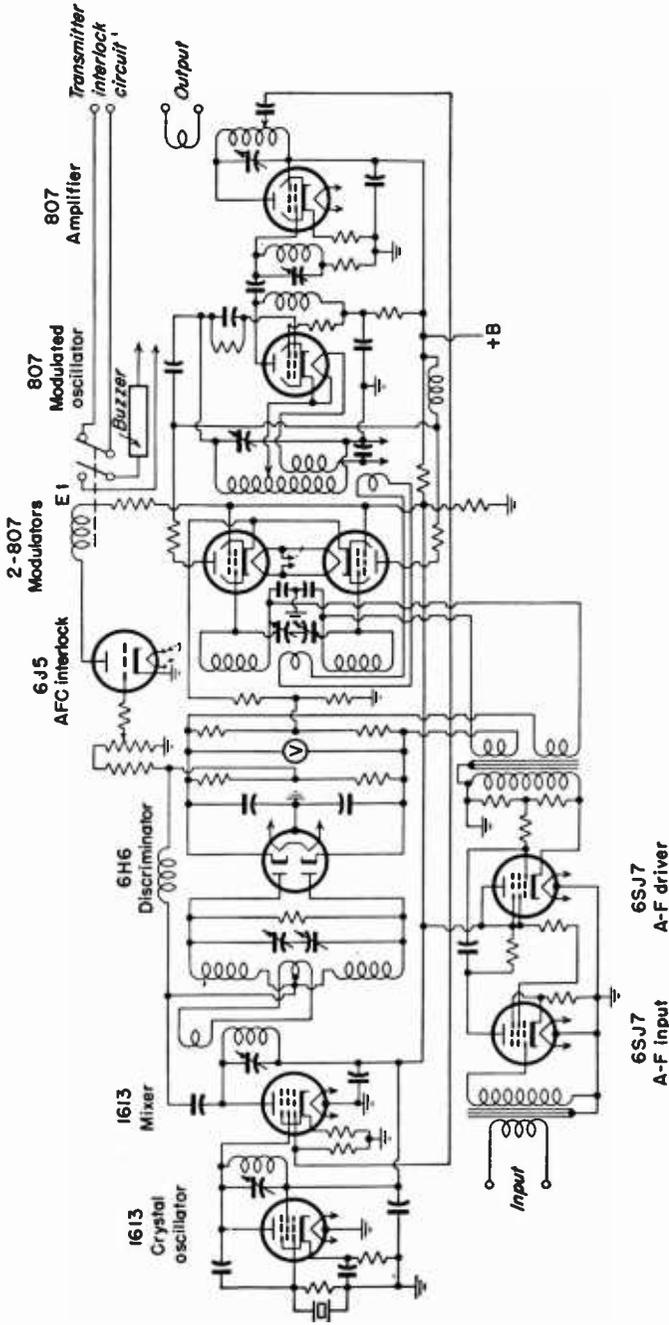
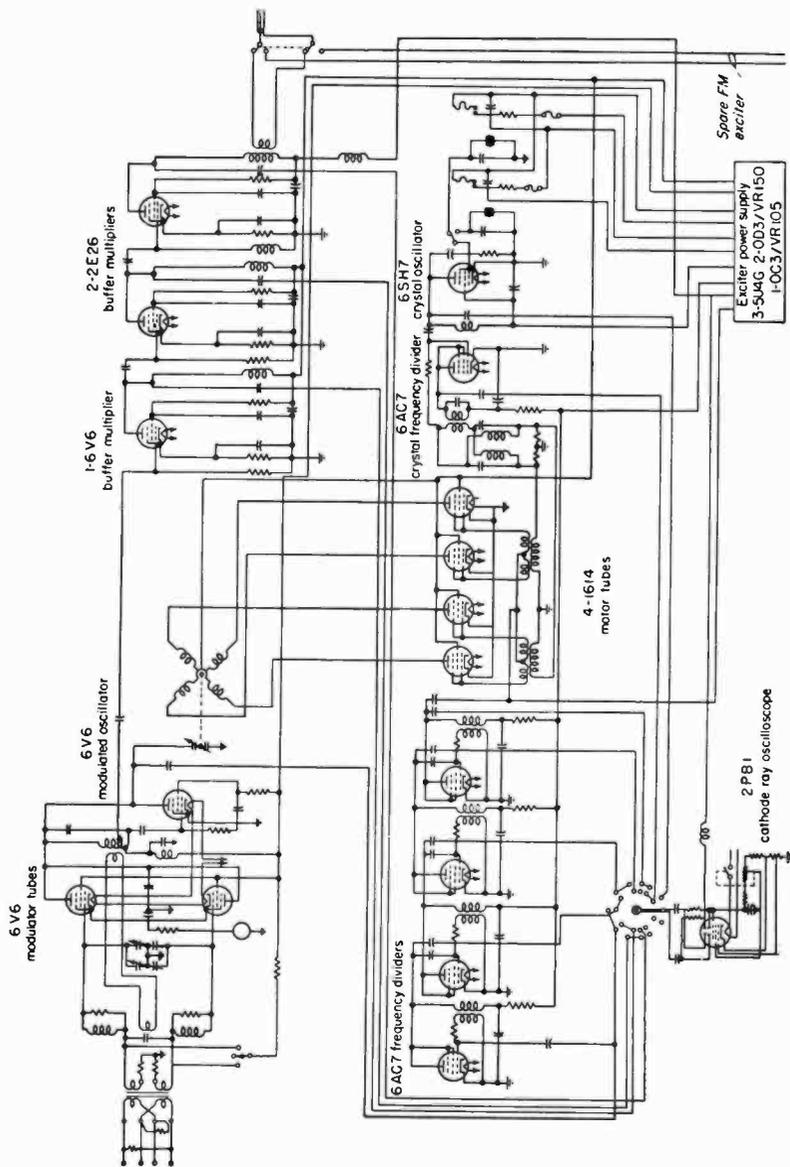


Fig. 68. Complete f-m transmitter circuit producing a compensating voltage from a frequency deviation.



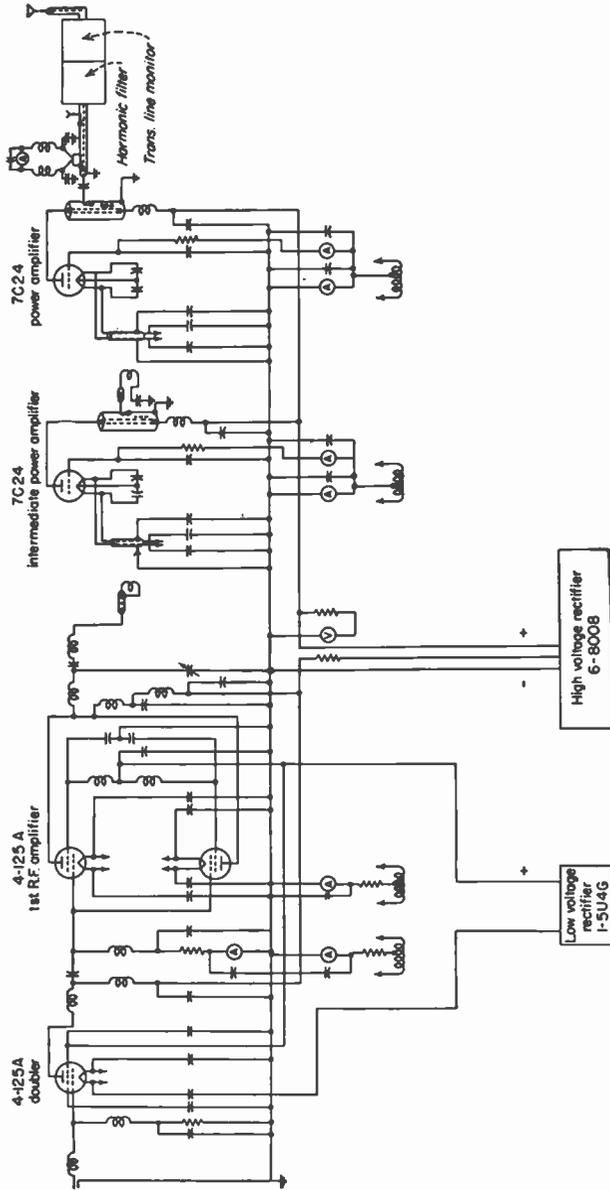


Fig. 69. The essential circuits of the output stages of the RCA 3-kw f-m transmitter.

Since the signal from the master oscillator provides a signal of the same phase for both modulators, then the 90-deg phase displacement in crystal-oscillator output signals supplied to the modulators will provide a similar phase displacement in the beat-frequency output from the two modulators. In this manner two-phase energy is available for driving a tuning motor to compensate frequency shift of the master oscillator. The direction of rotation of the 90-deg vectors representing the signals at the outputs of the two balanced modulators changes from clockwise to counterclockwise depending upon whether the master-oscillator frequency is higher or lower than the crystal standard frequency.

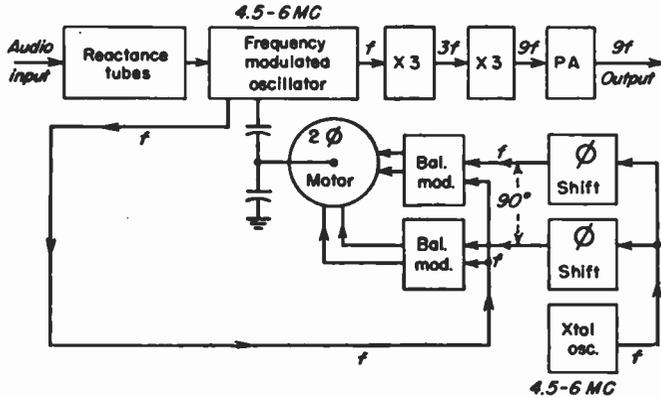


Fig. 70. Frequency-control f-m exciter.

The two-phase a-c output of the two modulators, therefore, has the characteristic of reversing one phase when the output secured from the master oscillator swings through zero beat from a higher to a lower frequency compared to the crystal.

The two-phase a-c output of the two modulators thus secured may be used to energize the windings of a two-phase motor. If the master-oscillator frequency is higher than that of the crystal, the motor shaft will rotate clockwise and so on. A compensating variable tuning capacitor may be attached to the motor shaft and employed to retune the master oscillator to zero beat with the crystal which operates on a submultiple of the assigned carrier frequency.

Circuits employed in the RCA BCF-3B reactance-tube modulator f-m system are shown in Figs. 69 and 70. The operating characteristics of this transmitter follow:

Frequency range.....	Any specified frequency between 88 and 108 Mc
Power output (into transmission line).....	1,000-3,000 watts
R-f output impedance.....	35-75 ohms
Carrier-frequency stability, deviation less than.....	2,000 cycles
Modulation capability.....	± 100 kc
Method of modulation.....	Reactance tubes
Audio input impedance.....	600 ohms
Avg program level.....	+4 ± 2 vu
100% modulation level.....	+12 ± 2 dbm
A-f response*	
30-15,000 cycles, uniform within.....	± 1 db
A-f distortion†	
30-100 cycles.....	1.5%
100-7,500 cycles.....	1.0%
7,500-15,000 cycles including all harmonics up to 30 kc	
at 75-kc swing.....	1.5%
F-m noise level, below ± 75-kc swing†.....	65 db
A-m noise level, below 100% amplitude modulation†.....	50 db

* For preemphasized response the preemphasis filter is provided to be inserted in the 600-ohm audio-input line at the most effective point.

† Distortion and noise are measured following a standard deemphasis network.

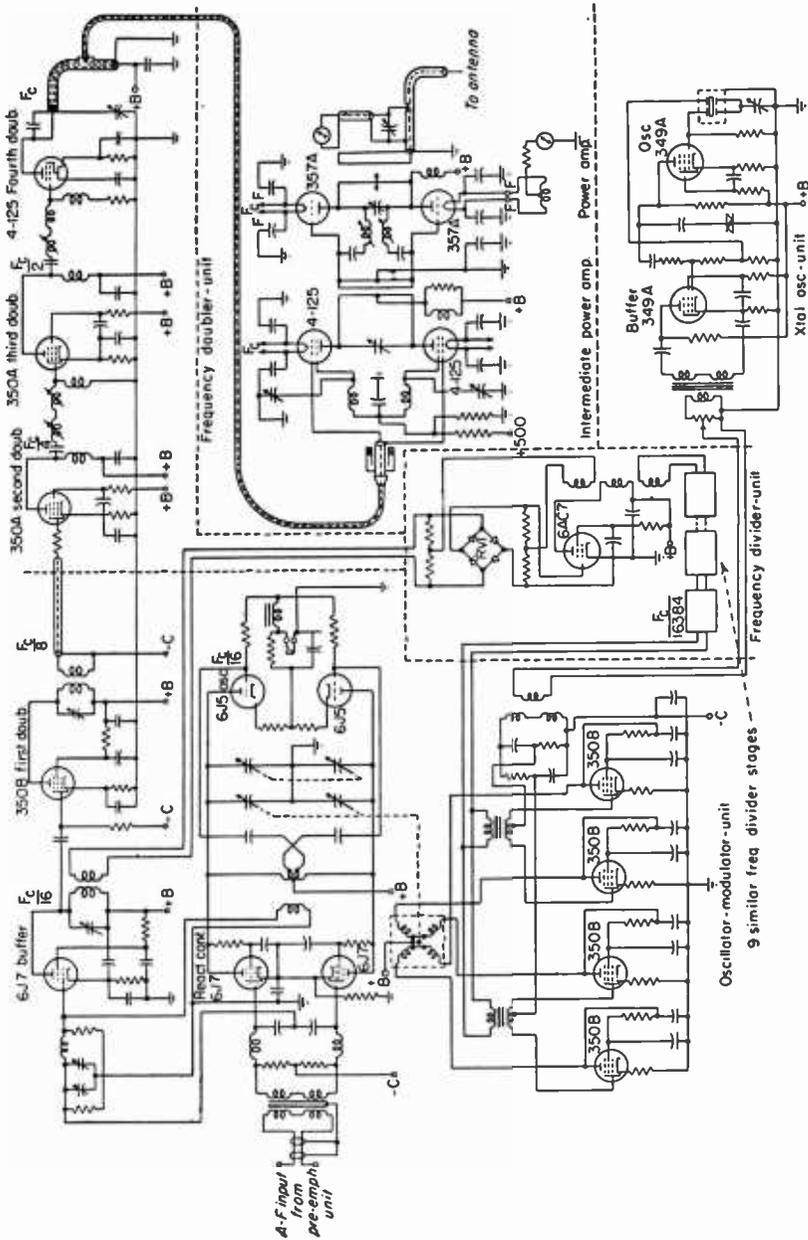


Fig. 71. The essential circuits of the 1-kw Western Electric f-m transmitter. This basic unit is used in all equipment of higher output power.

It is worth attention that grounded-grid amplifiers are used in the power amplifier and the driver. In this case the grid is maintained at ground potential, drive is applied between cathode and ground, and output is taken from grid and plate.

The 1-kw Western Electric type 503B-2 1-kw f-m¹ transmitter utilizes reactance tubes directly as frequency modulators in a manner as shown in Fig. 71. The method used to maintain constant the mean carrier frequency is called "synchronous f.m." since it operates by comparing the mean carrier frequency (measured in total carrier cps) of the f-m oscillator to a precision fixed-frequency standard. The difference frequency

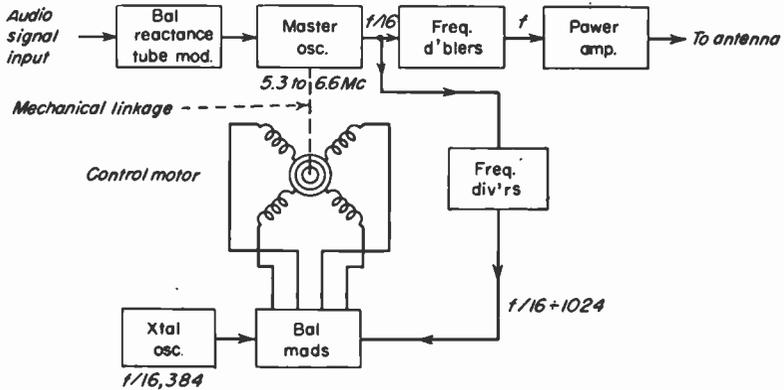


FIG. 72. Frequency-stabilizing system of Western Electric.

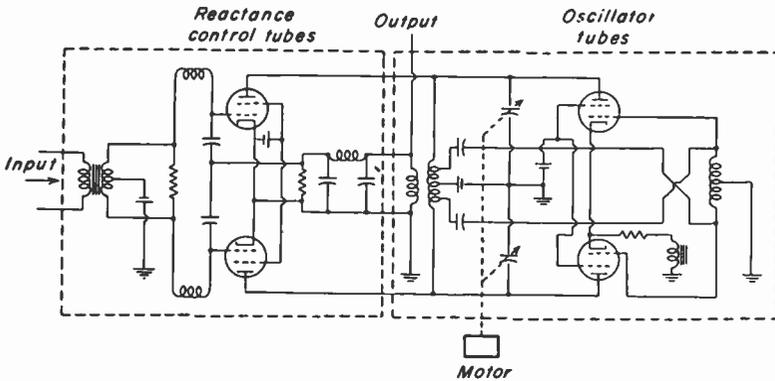


FIG. 73. Reactance tube circuit of Western Electric f-m transmitter.

thus derived is then utilized as a control medium for mechanically retuning the oscillator stage, keeping the oscillator frequency an exact multiple of the standard.

Figure 72 shows the frequency-stabilizing system, which functions through a small portion of the 6-Mc f-m oscillator (assume a 96-Mc carrier), being fed back through frequency dividers to obtain a 6-kc frequency equal to that of the precision quartz-crystal frequency standard. The 6-kc frequency, a much lower submultiple of the 96-Mc carrier, is necessary to produce a difference-frequency sufficiently low to be within the range of the rotating magnetic field of the electric motor used for retuning the oscillator variable capacitors through a suitable speed-reduction mechanism of gear

¹ Ganzenuher, J. H., Western Electric FM Broadcast Transmitters, *FM & Television*, September and October, 1946.

trains. The direction of rotation of the motor depends upon whether the oscillator frequency is higher or lower than the fixed multiple frequency of the standard. Thus automatic readjustment of the oscillator tuning is made in the correct direction so that when exact synchronism occurs between the frequency fed back from the f-m oscillator and the frequency standard the motor is at rest.

Because of the inertia of the motor rotating elements and the high order of frequency division used, the motor is not caused to rotate by frequency deviations produced on the carrier at an audio rate during modulation. The main advantage claimed for the synchronized f-m method is that the output frequency is maintained identical in precision to the standard by making all the controlling factors in terms of frequency.

63. Armstrong F-m Method. The indirect or p-m method of producing f.m.¹ consists in general of a constant-frequency oscillator, a modulator (the function of which is to change the phase of the oscillator output as illustrated in Fig. 74), and a series of multipliers to increase the amount of p.m. sufficiently to secure the frequency shift or modulation required in the radiated signal. Results are secured by splitting the oscillator output into two paths. One path contains a phase-shifting network which shifts the phase 90 deg and, in the other, a balanced modulator generating side bands with a suppressed carrier.

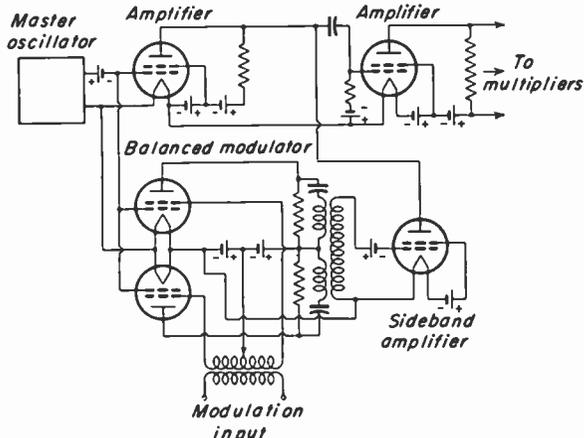


FIG. 74. Armstrong frequency modulator.

A combination of these two signals produces a p-m signal with a phase-shift modulation capability up to ± 30 deg with satisfactory linearity. An f-m wave is derived therefrom by transmitting the signal through frequency multipliers. A multiplication of several thousand times is required to obtain deviations of ± 75 kc.

To produce f.m. and at the same time maintain a constant deviation frequency, the p.m. must necessarily be inversely proportional to the modulating frequency. Therefore with this method it becomes necessary to have the amplitude of the p-m signal decrease in proportion to the frequency of the audio input to secure a flat a-f transmitter response. This is usually accomplished by a corrective network in the audio circuits ahead of the modulator.

The amount of frequency multiplication required following the production of phase modulation to secure the desired f.m. by the indirect method depends upon (1) the amount of p.m. produced by the modulator, (2) the lowest a.f. transmitted, and (3) the deviation or frequency swing at the output frequency. For a phase shift of 0.5 radian, frequency deviation of ± 60 kc and lowest a.f. 30 cps, the frequency multiplication required is 4,000 times.

To obtain this amount of frequency multiplication, the initial oscillator frequency must be multiplied in several stages, then heterodyned down to a lower frequency, and again multiplied a number of times to secure the output frequency.

64. Merits of F.M. vs. A.M. With the application of f.m. to transmitters operating in the v-h-f band, the relative merit for this system of signal transmission can be

¹ Armstrong, E. H., A Method of Reducing Disturbances in Radio Signaling by a System of Frequency Modulation, *Proc. IRE*, May, 1936.

evaluated on the basis of an a-m system. The v-h-f signal field intensity¹ at a given distance from a particular transmitting antenna may be determined from theoretical and empirical relationships as published in papers² and derived from extensive mathematical and experimental work. Actual experimental tests³ have shown that an interfering audio signal (output of receiver) will create objectionable interference if its level is about 30 to 40 db below the desired signal. Thus service areas can be defined as zones in which the desired component of the resulting audio signal is at least 35 db above the interference. For very high quality reproduction, this figure runs from 40 to 55 db.

For interfering signals on the same channel as the desired signal, it is evident that, if a.m. is used, a signal input ratio of 35 db is required to secure the desired output ratio. However, in f.m. the ratio of signals at the receiver input needs to be only about 6 db since the receiver for f-m reception responds to frequency variations and limits amplitude variations such as those caused by noise and undesired signals.

On this basis there are claimed advantages of f.m. over a.m. because of (1) improved signal-plus-noise to noise ratio. Experimental results have shown this difference to be as much as 25 db as influenced by intensities of automobile ignition, X rays, and other man-made interference. Atmospheric interference being small at ultrahigh frequencies, it becomes negligible in comparison with man-made interference. (2) A uniform and definite service area from a given transmitter since f-m signal-plus-noise to noise ratio remains high until field intensity reaches a low value. (3) A smaller geographical interference area obtained when two f-m transmitters are operated simultaneously on the same frequency as compared to similar operation of two a-m transmitters. (4) A r-f amplifier used to increase a f-m signal is more efficient than one used for a.m. because f.m. can be accomplished at low level followed by a class C r-f power amplifier. (5) For a given service area, less radiated power is required for f.m. because of the improvement in signal-plus-noise to noise ratio obtained with f.m. (6) For a given power output, power-tube operating costs are less because relatively smaller tubes can be used for f.m. for a given power output.

The FCC has provided for 65 channels each 200 kc wide for f-m carrier frequency assignments between 88 and 108 mc.

Preemphasis. A-f preemphasis at the transmitter is added as a means of increasing the effective modulation at the higher audio frequencies since, for average programs, the peak levels of modulation at these frequencies are lower than for the lower frequencies. At the receiver a corresponding amount of deemphasis is employed to bring the modulation back to its proper level. In this manner an increase in the signal-to-noise ratio is brought about.

Preemphasis is effective only when the inherent noise level of the transmitter at the point of preemphasis is low, since any noise or higher frequency disturbances already existing would be emphasized along with the microphone modulation. Since f-m transmitters generally have less noise (as required by the FCC) than a-m transmitters, this form of enhancing the higher audio frequencies is more effective with f.m. than with a.m.

Preemphasis is generally expressed in terms of the time constant of an *LR* circuit which produces the effect. The required value is that corresponding to a time constant of 75 μ sec.

An important advantage of the f-m transmitter over one utilizing the amplitude system lies in the fact that the efficiency of the r-f amplifier stages can be as great as it is for class C telegraph service. At ultrahigh frequencies this may be between 50 and 70 per cent.

Frequency-modulated Transmitter Measurements. The measuring equipment is considerably different than is required for an a-m station since there is a variation in fre-

¹ Trevor and Carter, Notes on Propagation of Waves below Ten Meters, *Proc. IRE*, March, 1933.

² Devino and Hunt, Ultra Short Wave Propagation over Land Burrows, *Proc. IRE*, December, 1935.

³ Weir, I. R., Field Tests of Frequency and Amplitude Modulation with U-h-f Waves, *Gen. Elec. Rev.*, May, 1939; Crosby, M. G., The Service Range of Frequency Modulation, *RCA Rev.*, January, 1940.

quency of the emitted wave with modulation while the amplitude is kept more or less constant. This is exactly the reverse of a.m. where the carrier is varied in amplitude but maintained at a constant frequency. The frequency swing or deviation can be measured by applying sustained tone to the transmitter and then measuring the relative intensities of the carrier and the side frequencies present, the relative amplitudes of which correspond to the Bessel functions involved.¹

TRANSMITTER DETAILS

65. Air- and Water-cooled Tubes. For tubes of low power, artificial cooling during operation is usually not necessary, radiation into the air being sufficient. For the larger tubes, however, artificial cooling is usually accomplished by means of a circulating water system which causes a sheet of water to pass over the anode surface at very high velocity.

To restrict leakage of current from the anodes to the grounded pipes of the water system, connection is made between the anodes and the water system through a long length of coiled hose or porcelain tubing. This interposes, between the anode and ground, columns of water long enough to make the electrical resistance to ground very high; as much as 100 ft of coiled hose may be used, giving resistances of 0.5 up to several megohms.

In many cases distilled water is used, the water being maintained at a satisfactory temperature by an artificial cooler, since for economical reasons it is desirable that the same water be used indefinitely.

The water-cooling and circulating system is automatically started when the transmitter is turned on, and the transmitter is automatically turned off in the event of any failure in the water-cooling system. One method of doing this is shown in Fig. 75, where the water system contains a Venturi tube whose inlet and output orifices are connected to a device operated by the difference in pressure established between the two orifices by the flow of water. If the flow is interrupted or falls below its normal value, a contactor through additional relays causes the power supply to be disconnected.

Sometimes a milliammeter is provided on the transmitter panel which indicates the magnitude of the current leaking through one of the closed coils, the amount of current serving to indicate the relative purity of the water and indicating when it is advisable to change the water supply.

In place of water cooling, forced air cooling is also used on some large tubes. For the large dissipation required, a large number of radiating fins are made a part of a copper radiator attached to the copper anode. Sufficient air is forced upward and between the cooling fins to carry away the heat developed on the anode. Because of the high electrostatic capacitance created by these anodes, they are not used on the very high frequencies with transmitters of very high power.

66. Power Supply. Plate-voltage supply for transmitters may be obtained from d-c generators, high-vacuum tube rectifiers, mercury-arc rectifiers, or hot-cathode mercury-vapor rectifiers.

The hot-cathode mercury-vapor rectifier is considered the best method of supplying high voltages to transmitter plate circuits. The most striking difference between mercury-vapor tubes and high-vacuum tubes is the internal voltage drop between plate and cathode. In the high-vacuum tube the voltage drop may vary from a few volts

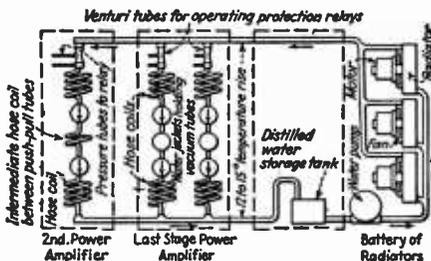


FIG. 75. Water-cooling and circulation system.

¹ Crosby, M. G., A Method of Measuring Frequency Deviation, RCA Rev., April, 1940.

to several thousand volts, depending upon the current, element spacing, etc. In the mercury-vapor tube the space charge is limited by the arc drop of the vapor which is practically constant at values between 12 and 17 volts regardless of the current.

Table 2 gives a direct comparison of the relative efficiency of a high-vacuum tube and a mercury-vapor tube. Note that the mercury-vapor tubes give very low internal voltage drop and have considerably higher efficiencies.

Two fundamental limits determine the power output that can be obtained from any number of tubes operated in any type of circuit: (1) the maximum peak inverse voltage at which the tube can operate without flashing back and (2) the maximum peak plate current which the cathode can supply with a reasonably long life.

The maximum peak inverse voltage which can exist across a tube in any of the usual types of circuits is equal to the line-to-line peak or crest voltage of the power transformer less the voltage drop of the conducting tube.

Table 2. Comparison of High-vacuum and Mercury-vapor Tube Rectifiers*

No. of tubes	Tube type	Circuit	D-c output			Tube drop		Losses, kw		Efficiency, per cent
			Volts	Amp	Kw	Volts	At amp	Fila-ment	Tube drop	
6	UV-214	3φ, double-Y	15,000	12	180	1,560	6	6.9	18.7	87.5
6	UV-857	3φ, full-wave	15,000	12	180	15	12	1.5	0.36	98.8
†6	UV-857	3φ, full-wave	21,000	30	630	15	30	1.5	0.9	99.4

* Proc. IRE, 18 (1), January, 1930.

† Maximum rating.

The peak plate current depends upon the type of circuit, tube, filter, and load. In a single-phase full-wave circuit each tube must carry the full-load current for half the time.

In the three-phase half- and full-wave circuit each tube carries the load current for one-third of the time. If the rectifier feeds into an inductance, square blocks of current are drawn from the rectifier and the peak plate current approaches the d-c value. If the rectifier feeds into a capacitance load plate current is drawn for only a part of each half cycle and the peak current may reach values of from three to five times that of the d-c load current.

Table 3 gives data on several typical hot-cathode mercury-vapor tubes designed for radio power supply purposes. The circuits most commonly used with these types of tubes are shown in Fig. 76. The single-phase full-wave and the three-phase and half-wave circuits are quite generally used. The three-phase full-wave circuit is particularly applicable to the half-wave mercury-vapor tube, since it gives a peak inverse voltage whose magnitude is only 4.5 per cent greater than the average output voltage; the wave form is that of a six-phase rectifier.

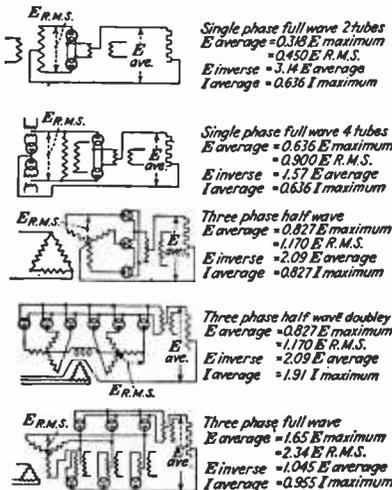


FIG. 76. Hot-cathode mercury-vapor power circuits.

67. Parasitic Oscillations. One of the most important design features of a transmitter is to provide for adequate suppression of parasitic oscillations. Such spurious

oscillations are usually caused by regeneration in an amplifier stage. They have frequencies different from the fundamental or its harmonics.

Table 3. Hot-cathode Mercury-vapor Tube Ratings

Tube type	Filament		Peak inverse voltage	Peak anode current, amp
	Volts	Amp		
UX-866	2.5	5	7,500	1.0
UV-872	5	10	7,500	5.0
UV-869A	5	18	20,000	10.0
UV-857B	5	30	22,000	40.0

All classes of amplifiers are subject to these oscillations. Suppressing them in a class C amplifier is not usually so difficult as in the class B types where the grids of the tubes are driven positive for a considerable portion of the cycle. Before reliable and economical service can be realized from a transmitter of any type, all tendencies for parasitic oscillation must be suppressed to prevent serious lessening in the life of vacuum tubes or program interruptions because of arc-overs in the transmitter. Such oscillations may exist in an otherwise normal amplifier stage and may not be evident to casual inspection owing to their disappearance entirely when grid excitation is removed.

A typical class B power amplifier stage of the push-pull type is shown in Fig. 77. This amplifier contains inherent design features which have a tendency to suppress spurious oscillations. C_6 and C_7 assist by acting as a very low reactance path for all parasitics of a frequency higher than the fundamental with a result that they effectively load the parasitic circuit. Connections between these capacitors and the tube grids are kept at an absolute minimum. The grid loading resistors R_1 and R_2 , the real purpose of which is to improve the regulation of the grid circuit as the grids swing positive, also act as a resistor load to damp out oscillations. C_8 and C_9 , with their mid-point grounded, act as a low-reactance path to ground for frequencies above the fundamental.

The frequency of parasitic oscillations may be anything from the very low end of the frequency spectrum to the u-h-f region. Parasitics of very low frequencies, in the neighborhood of less than 1 to 10 cycles, are sometimes set up by the dynatron action of the tubes at the natural period of the power-supply filter circuit C_1 , C_2 , and L_1 .

The existence of these parasitics of very low frequencies usually becomes apparent in the form of a severe irregularity in the saturation curve of the linear amplifier. Such a curve is shown in Fig. 78. The point X shows the beginning of this parasitic condition and Y the point where it ceases. It is caused by the dynatron characteristics of the amplifier tube grids and occurs at a point on their operating characteristic just before they are driven positive. A solution for such a condition is to use tubes whose amplification factor is such that the region XY falls below the carrier operating point. For this reason high- μ tubes have on some occasions been found to be more satisfactory than low- μ tubes.

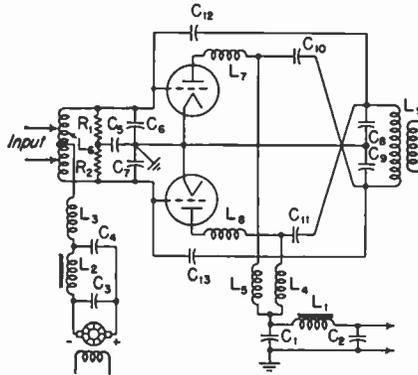


FIG. 77. Class B amplifier arranged for suppression of parasitics.

Low-frequency oscillations of approximately one-third to one-fifth of the fundamental frequency are sometimes caused by tuned-grid tuned-plate regeneration with the plate chokes L_5 and L_4 in combination with the blocking condensers C_{10} and C_{11} forming

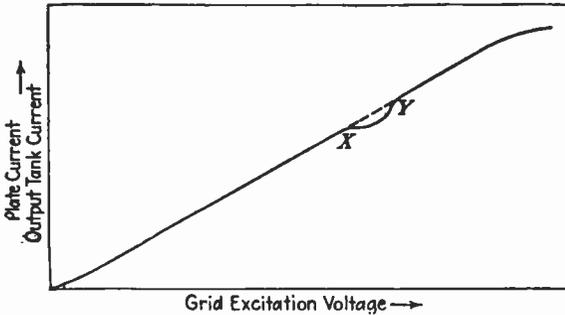


FIG. 78. Typical saturation curve of class B r-f linear amplifier showing dynatron effect of power tube grids with E_b and E_c constant.

ing an output tank circuit. A similar grid tank circuit is formed by C_6 , C_7 , and L_3 . Inasmuch as all tubes are effectively in parallel for this combination, the neutralizing capacitors tend to aggravate the condition rather than to prevent it. In Fig. 79 is shown an equivalent parasitic circuit of the combination as formed from the circuit in Fig. 77. The remedy is to change the values of inductance and capacitance in either the parasitic grid or plate circuits so as to cause their natural periods to depart substantially from a near resonance condition. It is usually possible to suppress such oscillation by tuning the parasitic grid circuit to a higher frequency than the corresponding plate circuit.

The existence of these oscillations may usually be detected by applying excitation at the fundamental frequency to a stage with reduced plate voltage and grid-bias voltage until the tubes draw plate current. If oscillation of the stage continues after fundamental grid excitation is removed, as indicated by neon lamps attached to the tube plates, the frequency of the parasitic may be determined by means of a wave-meter, and thus steps can be taken to eliminate it.

Oscillations within an amplifier stage at frequencies near the fundamental are usually caused by regeneration within an amplifier stage due to improper neutralization causing tuned-grid tuned-plate circuit oscillations. Improper circuit design or too close coupling between the inductances of the input and output circuits or chokes is also liable to cause this condition.

Parasitics of frequencies in the neighborhood of from 5 to 20 times the fundamental result in cases where the leads from the tube grids and C_6 and C_7 form a grid tank circuit, the resonance frequency of which is determined by various distributed capacitance and the inductance of the leads. Oscillations are made possible by the existence of a similar plate tank circuit formed by leads from the tube plates to C_9 and C_8 together with various stray capacitance. This form of parasitic is seldom sustained but shows itself most prominently when the stage is subject to high peaks of modulation. The trouble may usually be corrected by insertion in the plate leads at a point adjacent to the tube plates choke coils L_7 and L_8 .

These parasitic choke coils L_7 and L_8 together with a shortening of grid leads to an absolute minimum may also assist in suppressing oscillations of ultrahigh frequencies

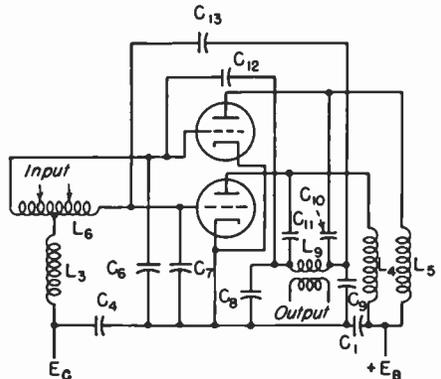


FIG. 79. Equivalent parasitic circuit of Fig. 77.

in amplifier stages employing two tubes in parallel. The grid leads of the two tubes, although connected, may combine with stray capacitance, thus forming a push-pull oscillation of a very high frequency. Such oscillations in some cases cause high r-f voltages to build up which may result in serious arc-overs from various parts of the tube output circuits.

68. **Suppression of R-f Harmonics.** It is the inherent characteristic of a vacuum tube, while functioning at a reasonably high efficiency in an amplifier circuit, to generate harmonic frequencies of the fundamental. A station broadcasting on 600 kc, if second and third harmonics were not suppressed, would produce interference with other stations operating on 1,200 and 1,800 kc. Field intensity measurements about a station are necessary to determine how much harmonic energy is radiated and to show the progress of work done toward reducing radiation.

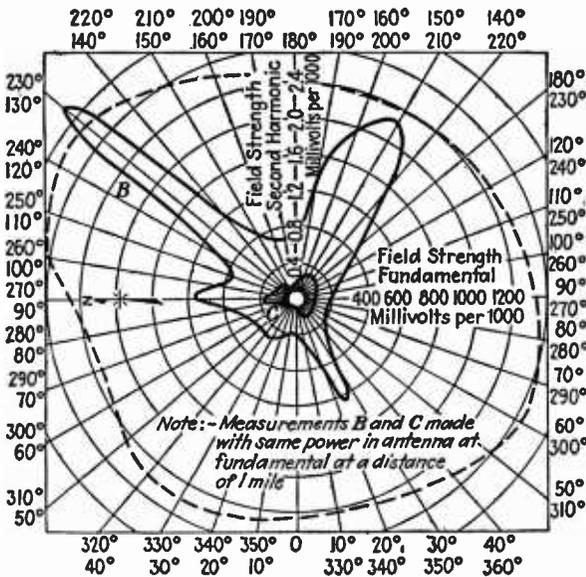


FIG. 80. Radio field-intensity survey. The dotted curve gives fundamental frequency field strength; B and C are second harmonic intensity before and after reduction.

In specifying the allowable harmonic radiation from a broadcasting station the IRE Committee on Broadcasting as of January, 1930, recommended that the maximum radio field intensity of a harmonic component measured at a distance of 1 mile from a station should not exceed 0.05 per cent of the field intensity of the fundamental. The FCC regulations prohibit radiation of a spurious or harmonic nature which may cause interference with stations operating on an assigned frequency.

A field strength, say, 500 μv per m at a distance of 1 mile is recommended as a maximum allowable intensity from a high-powered transmitting station. If in the case of a 50-kw station a circular-field pattern and equal attenuation are assumed for both a harmonic and the fundamental in the immediate vicinity of the station, a field strength of 500 μv at 1 mile would correspond to approximately 7 mw of radiated power at a harmonic frequency. The effect of directivity (illustrated in curve B, Fig. 80) may cause a field intensity of a number of times the value of 500 μv to be projected in a given direction with a very small fraction of 1 watt of harmonic power in the transmission line and antenna circuits. Such a concentration of radiated power may form very objectionable interference. Considering the factors involved, therefore, it is

evident that harmonic suppression must be attacked from a number of angles. These may be briefly outlined as follows:

1. Design of the transmitter circuits to reduce the harmonic content of the power delivered to the antenna circuits to a minimum.
2. Thorough and effective shielding of the entire transmitter or building.
3. Effectively grounding all harmonic drain circuits and elimination of long conductors near the transmitter coupled to it inductively or capacitively.
4. Reduction of directivity of harmonic radiation to a minimum.
5. Installation of shielded band- or low-pass filters at the input end of the transmission line to the antenna.

Some commonly used triode amplifier circuits are shown in Figs. 81 and 82. The push-pull amplifier is superior to the single-ended circuit, as it is capable of producing

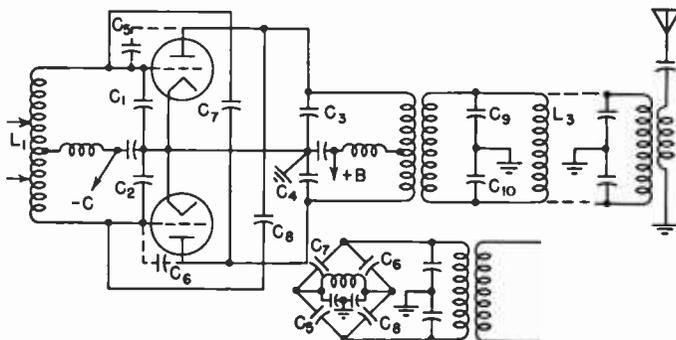


FIG. 81. Push-pull amplifier with high kva tank circuit in transmission line.

a sum plate current of the two tubes which is symmetrical in wave shape and, therefore, it contains no even harmonics. Individual plate currents, of course, contain even harmonics which are drained to ground through C_3 and C_4 resulting in identical instantaneous even-harmonic potentials being set up on each side of L_2 but no actual even-harmonic current through it. Under these conditions an electrostatically shielded inductive coupling is provided to permit transfer of only fundamental and odd-harmonic frequencies to the coupled circuit. For a condition of symmetrical plate current it is evident that the tube characteristics must match closely, $C_1 = C_2$ and $C_3 = C_4$. The neutralizing bridge must be balanced not only for the fundamental frequency but for even harmonics. This requires that the internal capacitance of the tubes should match. As will be shown later, a high ratio of circulating kilovolt-ampere in the tank circuit to the kilowatt delivered from the amplifier reduces the output of harmonics from a single-ended amplifier to a very low value. This is also true in the push-pull circuit.

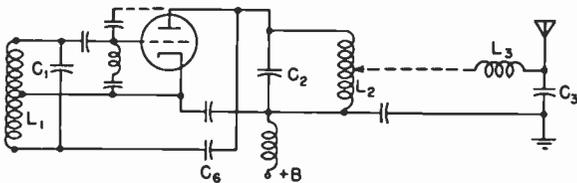


FIG. 82. Line termination effecting reduced harmonic radiation.

The circuit shown in Fig. 82 will give a very small amount of harmonic output by proper design of the circuit constants. The curves in Fig. 84 show the filtering effect of a high kilovolt-ampere tank circuit in suppressing harmonic components of current generated in the tube. These curves show actual harmonic transferred to a given load circuit Z_L with a constant output at the fundamental and various kilovolt-ampere to kilowatt ratios of L_2 and C_1 . Figure 83 shows improvement in tank circuits so as to

increase the normal filtering action of an ordinary tank circuit. A high kilovolt-ampere to kilowatt ratio applied to these circuits is capable of reducing harmonic output to an extremely small amount. There are some limitations in the amount of filtering which can be secured by a high kilovolt-ampere tank circuit, however, since the I^2R losses in the circuit increase in proportion to the circulating kilovolt-amperes and the cost of apparatus for increasing kilovolt-amperes in a circuit without increasing losses is considerable. In broadcasting transmitters there is the limitation of too low a decrement in a circuit attenuating too greatly the high frequencies of a modulated envelope. In Fig. 83 the trap L_3C_3 is tuned to a particular harmonic to be eliminated. The use of antiresonant circuits (parallel traps) in the plate lead of an amplifier, while reducing to some extent a single harmonic, has a tendency to allow considerable voltage to build up at others. Most satisfactory results are usually secured by designing a minimum impedance path for harmonics to ground as compared with a given high impedance at the fundamental.

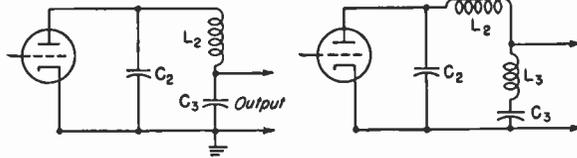


FIG. 83. Improved tank circuits for suppressing harmonic radiation.

The effectiveness of the shielding of a transmitter may be determined by operating the transmitter with full power output into a shielded phantom antenna. Measurement of the harmonic field strengths produced from the transmitter itself is direct evidence of how well it is shielded. Such radiation can usually be traced to a long conductor near the transmitter, coupled to it capacitively or through a common ground return. Ground conductors serving to drain harmonic frequency power to ground therefore should be as direct as possible and should not be extended so

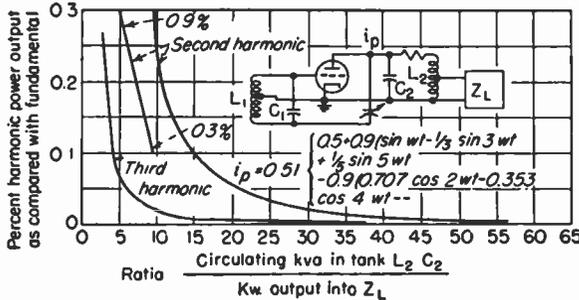


FIG. 84. Effectiveness of high kva to kw ratio in reducing harmonic output with constant power output at fundamental.

as to have a free end which might attain a high potential at resonant frequencies. This is particularly true of the harmonic drains near the antenna itself. These should have a separate ground to prevent coupling of harmonic frequencies into the antenna.

A sensitive wavemeter is very useful in determining the relative harmonic field intensities near the various circuits of a transmitter. When tuned to the frequencies of various harmonics and coupled to various circuits of the transmitter or placed at positions along nearby open conductors, this instrument will indicate proportionate amounts of the harmonic components of the current flow. By effectively grounding a long open conductor, either directly or through large capacitors at a number of distributed points, harmonic radiation can usually be eliminated.

The push-pull amplifier coupled to a long transmission line has often become a source of undesirable even-harmonic radiation because of sufficient electrostatic capacitance existing between the coupled circuits to permit a transfer of energy from the amplifier output circuit to the line. Unless this electrostatic capacitance is reduced to an extremely low value, e.g., by installation of a well-grounded electrostatic screen between the two coils, even harmonics usually find a path along the transmission line with a

ground return to the generating source. An unshielded transmission line serves in this case as an effective directive radiator in the form of a large loop. Its effective height will be dependent upon the height of the transmission line above ground. Parallel flow of even-harmonic currents along the line, therefore, makes it a much more effective radiator in some directions than the push-pull flow of harmonic currents in the line.

A circuit which has been found to be very effective in reducing both the parallel as well as the push-pull flow of harmonic currents in a transmission line is shown in Fig. 81 in the form of a high kilovolt-ampere floating tank circuit $L_3C_9C_{10}$ tuned to the fundamental component of current flowing in the line. This tank circuit, while offering an impedance to the fundamental approaching an infinitely high value, offers a relatively low impedance path to ground for the parallel flow of even harmonics equivalent to

$$Z_{ne} = \frac{-1}{4\pi f_{ne} C_9} = \frac{-1}{4\pi f_{ne} C_{10}}$$

where resistance of circuit is negligible, Z_{ne} = impedance to n th even harmonic, and f_{ne} = frequency of n th even harmonic, and for the push-pull flow of odd harmonics between transmission-line conductors

$$Z_{no} = \frac{-2\pi f_{no} L_3}{(2\pi f_{no})^2 L_3 C - 1}$$

where resistance of circuit is negligible, Z_{no} = impedance to n th odd harmonic, f_{no} = frequency of n th odd harmonic, and $C = C_9/2 = C_{10}/2$, where $C_9 = C_{10}$.

It is evident that as C_9 and C_{10} are increased in value the effectiveness of the circuit in reducing harmonics is increased. Since the transmission-line termination

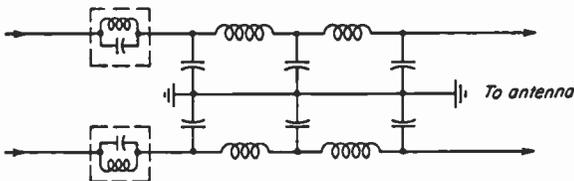


FIG. 85. Low-pass filter combined with antiresonant circuits in transmission line.

impedance is usually made to match the line impedance for the fundamental frequency, it usually happens that the line impedance is matched for this frequency only and as a result harmonic components of current and voltage in the line appear as standing waves along the line. In such a case the above tank circuit is most effective for eliminating a particular harmonic if it is placed at a point along the line of maximum voltage. This circuit alone was effective in one case in reducing second-harmonic radiation from a station to one-fifth of its former value.

Antiresonant circuits installed in a transmission line at current antinodes have been found very effective in reducing a single harmonic to which they were tuned. Extreme care should be taken in shielding these antiresonant circuits to secure best results. A combination of antiresonant circuits and a low-pass filter is shown in Fig. 85. This combination has been used successfully in severe cases of harmonic radiation from a very long transmission line and antenna system. The filter matches the surge impedance of the line and has a cutoff frequency between the fundamental and second harmonic. Antiresonant circuits have been found useful to sharpen the cutoff so as to attenuate sufficiently the second-harmonic frequency. Considerable experience in filter design and adjustment is required to secure optimum results from such an arrangement. For use with concentric lines with the outer sheath grounded, the filter shown in Fig. 85 is simplified to the extent of one-half, *i.e.*, one line to ground.

The methods of line termination shown in Figs. 81 and 82 are effective in reducing the possibility of harmonics reaching the antenna circuit. The termination shown in Fig. 82 may be improved by use of a multisection low-pass filter.

69. Antenna Circuit Terminations for R-f Transmission Lines. Considerable improvement in antenna efficiency can be secured from an antenna located at some distance from the station so as to approach the ideal case of an antenna radiating in free space. The r-f transmission line is used for conveying the energy from the trans-

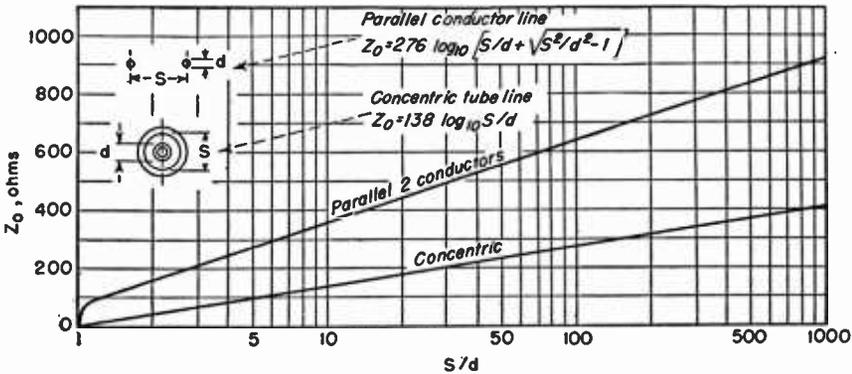


FIG. 86. Impedance of parallel and concentric-tube lines.

mitter to the antenna. A simple form of such a transmission line is the parallel two-conductor type, each conductor having a diameter of approximately $\frac{1}{2}$ in. The spacing of the conductors is normally 12 to 15 in.

The curves of Fig. 86 show the characteristic impedance values with respect to spacing and conductor size of both the parallel conductor line and the concentric-tube type.

70. Transmission-line Calculations. There are diverse methods of measuring the characteristic impedance of a transmission line. A simple but effective method is illustrated in Fig. 87. With the setup shown and the switch thrown to the line position, a trial value of resistance R_2 is inserted. C is adjusted for maximum I_1 . Then with switch thrown in the opposite position and R_1 set to equal R_2 , the capacitor C is adjusted for maximum I_1 . By trial, a combination may be found where there is a maximum value of I_1 and I_2 for the same setting of C with R_1 equal to R_2 . This value of R is the characteristic or surge impedance of the line.

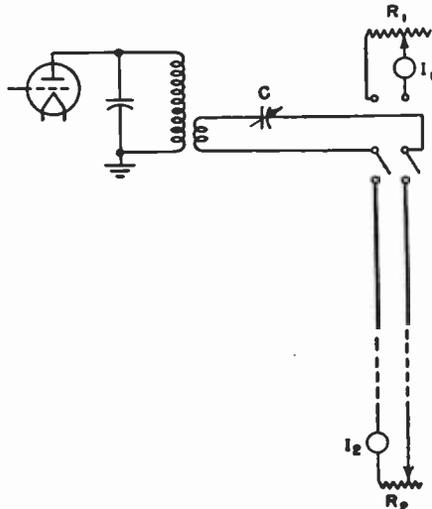


FIG. 87. Measuring impedance of transmission line.

When r-f power is transmitted over a transmission line to an antenna load, the line termination may be adjusted to afford a condition where there are no wave reflections by making the effective resistance of the termination equal to the characteristic impedance of the line.

Several circuits used for terminating transmission lines are shown in Figs. 88 to 90 together with their equivalent circuits.

A formula for calculating the value of capacitor C_B for an effective resistance value Z_0 equal to the characteristic impedance of a two-conductor transmission line balanced

to ground as shown in Fig. 90 as well as for a transmission line having one conductor grounded is as follows:

Let Z_0 = effective resistance of transmission-line termination

R_a = antenna resistance consisting of radiation resistance plus equivalent loss resistance

L_T = combined inductance-balance coils plus equivalent antenna inductance

C_A = equivalent antenna capacitance

C_B = line-termination capacitance

X_1 = reactance of C_B

X_2 = reactance of $X_L - X_{CA}$

Z_1 = impedance branch 1 = $-jX_1$

Z_2 = impedance branch 2 = $R_a + jX_2$

$$Z_0 = \frac{R_a X_1^2 - j(X_2^2 - X_1 X_2 + R_a^2)}{R_a^2 + (X_2 - X_1)^2}$$

$$C_B = \sqrt{\frac{Z_0 - R_a}{4\pi^2 f^2 Z_0^2 R_a}}$$

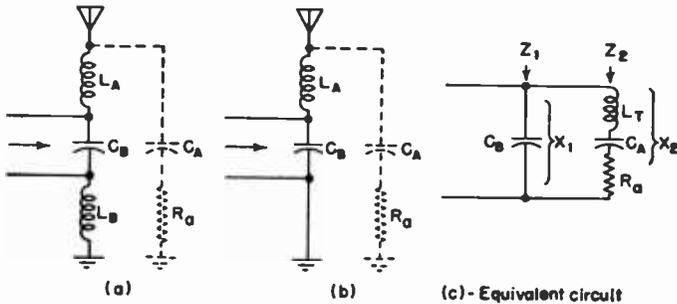


FIG. 88. Terminations for transmission lines.

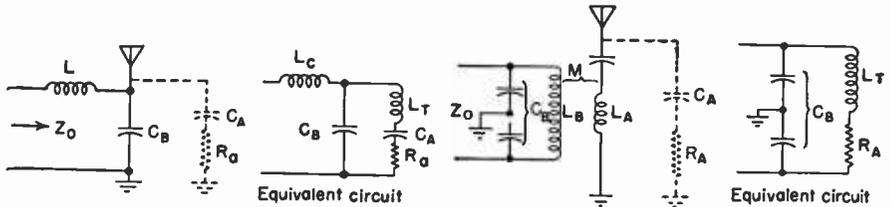


FIG. 89. Transmission-line termination.

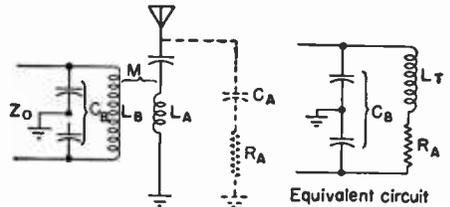


FIG. 90. Balanced transmission-line termination.

where $Z_0 > R_a$

$$X_1 = \frac{Z_0 R_a}{\pm \sqrt{R_a(Z_0 - R_a)}} = \frac{1}{2\pi f C_B}$$

where, in Fig. 90, C_B is dependent only on values of Z_0 and R_a where Z_0 is equivalent to a pure a-c resistance with the antenna circuit adjusted for resonance. Unless Z_0 exceeds the value of R_a , an effective resistance equivalent to the characteristic impedance of the line cannot be secured.

When low-impedance lines are used, such as the concentric-tube type, the termination shown in Fig. 89 is useful, since it affords a condition where correct termination may occur in the form of an effective resistance even though R_a equals or exceeds Z_0 .

In Fig. 90 is shown a transmission line termination in the form of a tank circuit. The tank capacitor C_B across the line is selected so as to provide a suitable kilovolt-ampere ratio of the tank circuit with respect to the kilowatts transferred to the antenna circuit; this kilovolt-ampere to kilowatt ratio is normally about 10 and should never be less than 2.

$$X_1 = \frac{Z_0 R_a}{\pm \sqrt{R_a(Z_0 - R_a)}}$$

$$X_1^2 = \frac{Z_0^2 R_a}{(Z_0 - R_a)}$$

from which

$$R_a = \frac{X_1^2 Z_0}{Z_0 + X_1^2}$$

where R_a is the effective value of resistance reflected into the tank circuit from the antenna circuit.

The value of R_a can be calculated from

$$R_a = \frac{\omega^2 M^2 R_a}{R_a^2 + X_a^2}$$

where the inherent resistance of the tank circuit is negligible, M = mutual inductance between L_A and L_B , X_a = reactance of antenna circuit, and R_a = resistance of antenna circuit.

For a condition of proper termination X_a approaches zero and may be neglected and

$$\frac{\omega^2 M^2}{R_a} = \frac{X_1^2 Z_0}{Z_0^2 + X_1^2}$$

$$M = \sqrt{\frac{X_1^2 Z_0 R_a}{\omega^2 (Z_0^2 + X_1^2)}}$$

In Fig. 91 are shown values of M required for a transmission impedance of 400, 500, and 600 ohms and a line-termination capacitor of between 0.001 and 0.004 μf . The transmitter frequency was assumed as 670 kc and the antenna resistance 30, 70, and 140 ohms. In the design of a tank-circuit termination for a given line the value of C_B across the line is selected so as to provide a proper kilovolt-ampere in the tank circuit with respect to the power transferred to the antenna circuit. This kilovolt-ampere to kilowatt ratio is normally about 10.

71. Termination Adjustments. The usual procedure in adjusting a transmission-line termination for a condition of no-wave reflection on the line is as follows:

1. The number of coupling turns is calculated so as to give the proper value of M . With the tank circuit open, the antenna is tuned to exact resonance by means of an external oscillator loosely coupled to it at the fundamental frequency.

2. The tank circuit is now connected into the circuit and tuned to resonance. This is indicated by a condition where the current in the antenna circuit becomes a minimum.

3. The transmission line is then connected across the tank circuit without making any changes in previous adjustments.

4. Correct termination may be checked by measuring the transmission-line currents at the ends and $\lambda/4$ points along the line by means of suitable meters. When proper termination has been effected, the transmission-line currents will be identical at all points along the line.

72. Concentric Line Terminations.¹ The growing use of concentric lines of the low-impedance type has led to cases where the characteristic impedance of the transmission

¹ This and the following section are from *Electronics*, December, 1936.

line is lower than that of the antenna resistance. In general there are three cases to consider as follows: (1) when the antenna impedance contains a resistance component only, (2) when the antenna impedance contains a resistance component and a reactive component, either (a) capacitive or (b) inductive, and (3) when the antenna impedance contains resistive and reactive components, the latter being partially compensated by

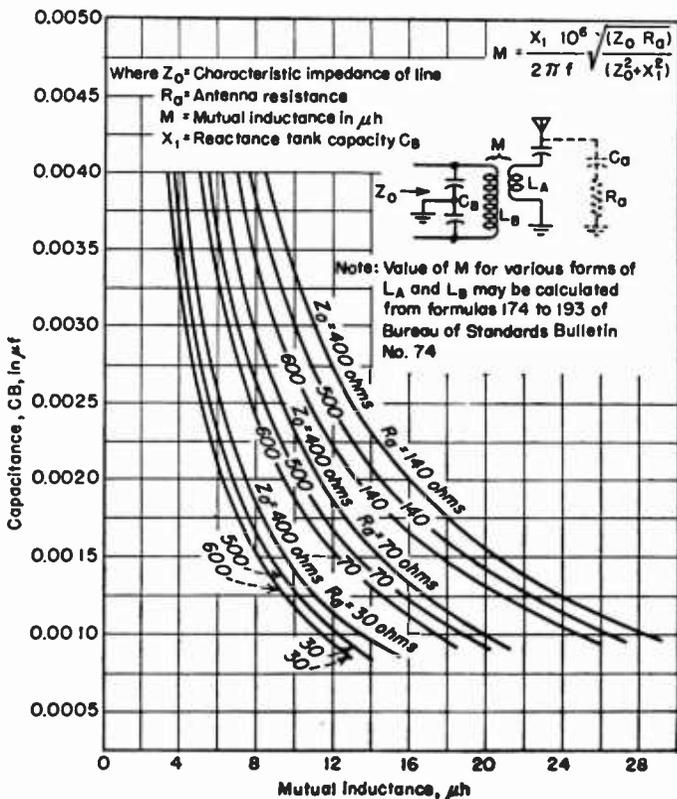


FIG. 91. Value of M required for proper termination.

the insertion of an extra reactance of opposite sign. These three cases are considered in order.

Case 1. Antenna Impedance Purely Resistive. From Fig. 92 the concentric line characteristic impedance, Z_0 , is terminated by a network consisting of C_B , L_C , and the antenna impedance Z_A . For case 1 the reactance of the antenna impedance is zero, and

$$Z_0 < Z_A = R_A.$$

Then the complex impedance Z_L presented to the end of the transmission line is as follows:

$$Z_L = \frac{R_A[X_1 X_1 - X_1(X_1 - X_1)] + j[X_1 X_1^2 + R_A(X_1 - X_1)]}{R_A^2 + X_1^2}$$

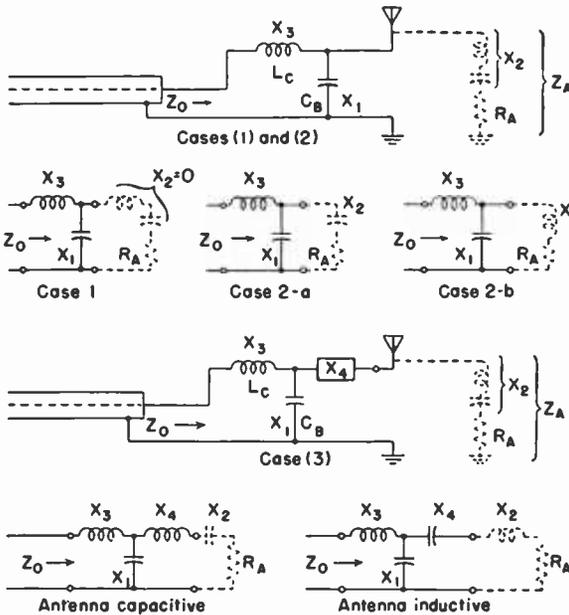


FIG. 92. Concentric line terminations.

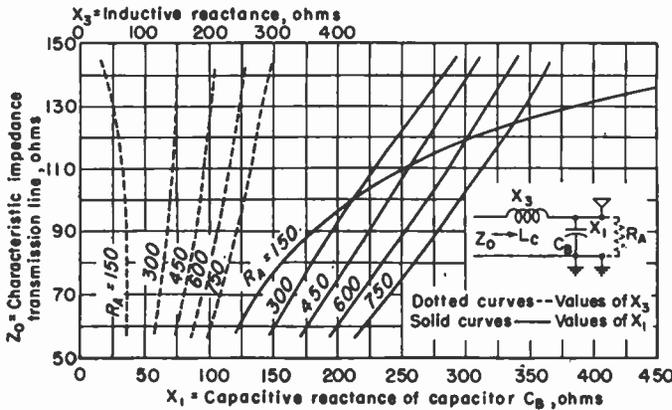


FIG. 93. Reactances for line terminations.

where R_A , X_1 , and X_3 are as given in Fig. 92. For proper termination Z_0 must equal Z_L . X_1 becomes

$$X_1 = R_A \sqrt{\frac{Z_0}{R_A - Z_0}}$$

$$X_3 = \frac{R_A^2 X_1}{X_1^2 + R_A^2}$$

Since $C_B = 1/(2\pi f X_1)$ and $L_C = X_3/(2\pi f)$, their values in microfarads and microhenrys are then readily calculable from f , the frequency of operation. Figure 93 gives various values of X_1 and X_3 in terms of values of Z_0 and R_A .

Case 2a. *Antenna Impedance with Capacitive Reactance.* Refer again to Fig. 92. It will be noted that the equivalent diagram for case 2a is the same as for case 1, except that the antenna impedance is now $Z_A = R_A - jX_2$. Then

$$Z_L = Z_0$$

$$Z_0 = \frac{R_A X_1^2}{R_A^2 + (X_2 + X_1)^2}$$

from which

$$X_1 = \frac{Z_0}{R_A - Z_0} \left[X_2 \pm \sqrt{\frac{R_A}{Z_0} (R_A^2 + X_2^2 - Z_0 R_A)} \right]$$

$$X_3 = \frac{X_1 (R_A^2 + X_2 X_1 + X_1^2)}{R_A^2 + (X_1 + X_2)^2}$$

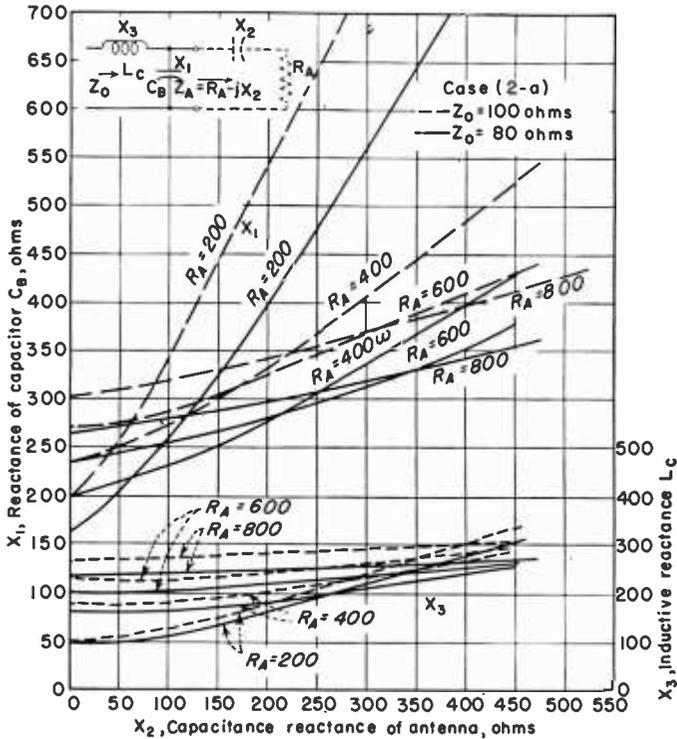


FIG. 94. Values of terminating reactors for case 2a.

Hence, with Z_0 , R_A , and X_2 given, X_1 and X_3 can be calculated. From the values of X_1 and X_3 , L_C and C_B can be calculated, exactly as in case 1. Values of X_1 and X_3 for various values of R_A and values of X_2 for the cases where Z_0 is 80 and 100 ohms are given in Fig. 94.

Case 2b. *Antenna Impedance Inductively Reactive.* Case 2b is the same as case 1 except that $Z_A = R_A + jX_2$.

$$X_1 = \frac{Z_0}{R_A - Z_0} \left[-X_2 \pm \sqrt{\frac{R_A}{Z_0} (R_A^2 + X_2^2 - Z_0 R_A)} \right]$$

$$X_3 = \frac{X_1 (R_A^2 + X_2^2 - X_2 X_1)}{R_A^2 + (X_2 - X_1)^2}$$

from which L_c and C_B are calculated. Figure 95 shows various values of X_1 and X_2 in terms of R_A and X_2 , for Z_0 values of 80 and 100 ohms.

Case 3. Added Reactance to Antenna Impedance. When the transmission line impedance "looks into" a complex antenna impedance, it is possible to simplify the adjustment of the circuit greatly by adding a reactance X_4 as shown in Fig. 92 for case 3. This reactance X_4 may be either inductive or capacitive, as shown. If the sum of X_2 and X_3 is inductive,

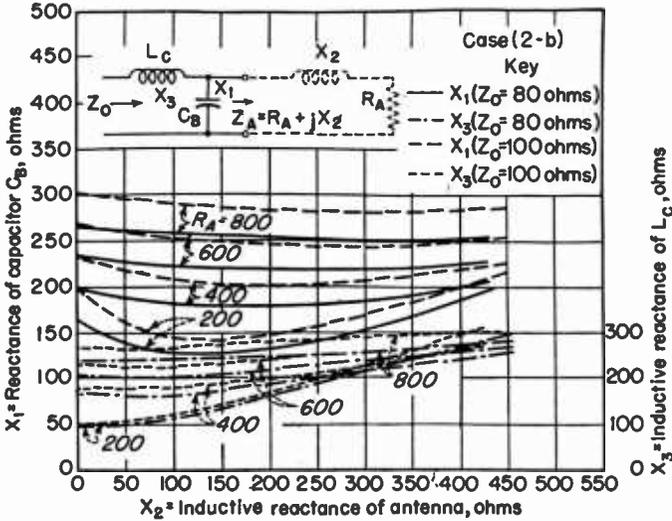


FIG. 95. Terminating reactors for case 2b.

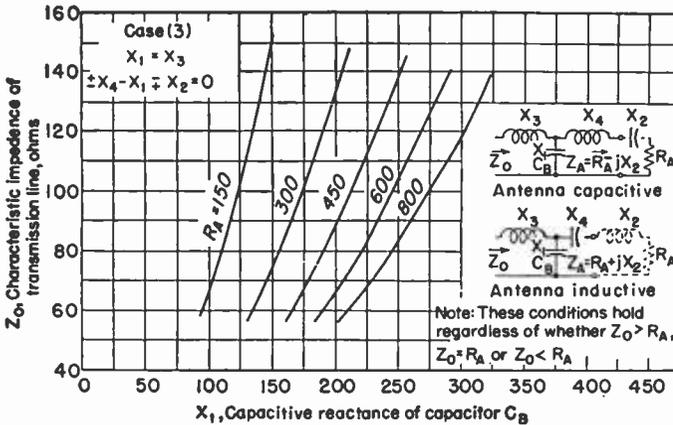


FIG. 96. Terminating reactors for case 3 when antenna reactance is compensated.

then X_4 is made capacitive, and vice versa. The value of X_4 is such that the algebraic sum of X_1 , X_2 , and X_4 is equal to zero. Since X_4 is in series with the antenna impedance, it adds directly with the reactive part of the antenna impedance. The effect of the presence of X_4 can then be taken into account of applying the formulas of case 2a or 2b.

$$X_1 = X_3 = \sqrt{Z_0 R_A}$$

This occurs only, however, if X_4 is so chosen that

$$\pm X_4 - X_1 \mp X_2 = 0$$

The reactance X_4 must always have the opposite sign from X_1 , as indicated by the plus-minus signs in the equation. When X_4 is so chosen, the reactance X_1 and X_2 may be obtained for various values of Z_0 and R_A by reference to Fig. 96. Note that these values apply regardless of whether R_A is larger than, equal to, or greater than Z_0 .

Practical Procedure in Designing Matching Circuits. In making suitable adjustments on the impedance matching circuits to provide a correct termination for a given

CURRENT AND VOLTAGE RELATIONS



CASE	ANTENNA CURRENT I_A	CURRENT THROUGH CAPACITOR $C_B = I_C \left(\frac{Z_A}{Z_A + Z_B} \right) = I_B$	TRANSMISSION LINE CURRENT $I_A \sqrt{\frac{R_A}{Z_0}} = I_C$	VOLTAGE AT BASE OF ANTENNA $E_A = I_A Z_A$
1	$\sqrt{\frac{W}{R_A}}$	$I_C \left[\frac{R_A^2 + jR_A X_1}{R_A^2 + X_1^2} \right]$	$\sqrt{\frac{W}{Z_0}}$	$I_A R_A$
2a	$\sqrt{\frac{W}{R_A}}$	$I_C \left[\frac{R_A^2 X_1 X_2 + X_2^2 jR_A X_1}{R_A^2 + (X_2 + X_1)^2} \right]$	$\sqrt{\frac{W}{Z_0}}$	$I_A (R_A + jX_2)$
2b	$\sqrt{\frac{W}{R_A}}$	$I_C \left[\frac{R_A^2 X_1 X_2 + X_2^2 jR_A X_1}{R_A^2 + (X_2 - X_1)^2} \right]$	$\sqrt{\frac{W}{Z_0}}$	$I_A (R_A - jX_2)$
3 ANTENNA CAPACITIVE	$\sqrt{\frac{W}{R_A}}$	$I_C \left[\frac{R_A + jX_1}{R_A} \right]$	$\sqrt{\frac{W}{Z_0}}$	$I_A (R_A - jX_2)$
3 ANTENNA INDUCTIVE	$\sqrt{\frac{W}{R_A}}$	$I_C \left[\frac{R_A + jX_1}{R_A} \right]$	$\sqrt{\frac{W}{Z_0}}$	$I_A (R_A + jX_2)$

W is power in watts

FIG. 97. Current and voltage relations in terminating circuits.

transmission-line characteristic impedance, under cases 2a and 2b above, where $R_A > Z_0$, the following procedure is recommended:

1. The transmission-line characteristic impedance should be calculated and the results checked by actual measurements if possible, either by means of an r-f impedance bridge or by the method described previously.
2. The antenna base resistance should be measured over a frequency band width covering at least 100 kc each side of the operating frequency. A curve should then be constructed with values of antenna resistance as a function of frequency. A smooth curve drawn through the points of measurements will assist in checking their accuracy.
3. Together with antenna resistance measurements, the antenna reactance should be measured, either by means of an r-f impedance bridge or in a manner shown in Fig. 98 over a wide frequency range and a curve constructed with antenna reactance as a function of frequency.
4. With the values of antenna resistance and reactance known, values of capacitance C_B and inductance L_C may be calculated for case 2a or 2b, as may be required, and connected into the circuits as shown in Fig. 92.
5. With the transmission line connected, correct termination may be checked by measuring the transmission-line currents at the ends, if its length is equal to $\lambda/4$ or odd multiples thereof. For a very long line it is good practice to make these measurements at a number of points along the line. The existence of stationary waves of current or voltage of the fundamental frequency along the line is an indication of incorrect termination. In such a

case slight adjustments may be necessary in L_C and L_B to correct for stray capacitance of leads and tuning equipment or slight errors in measurements. If an r-f impedance bridge is available, its measuring terminals may be connected across the input to the matching circuit in place of the transmission line and the termination circuit checked for an effective resistance equivalent to the characteristic impedance of the line without the line attached.

Although case 3 requires the addition of another piece of apparatus in the form of an inductance or capacitance in the antenna lead, which may be rather expensive, the adjustment procedure is less difficult and is as follows:

1. With values of the line characteristic impedance, antenna resistance, and reactance obtained by measurement, the value of C_B is calculated, which gives the reactance X_1 necessary.
2. With L_C disconnected from C_B , reactance X_4 (inductive or capacitive) is added in the antenna circuit in series with X_1 . By means of X_4 the antenna circuit is tuned to resonance, as indicated by maximum current through a thermogalvanometer, when the antenna circuit is excited by means of an external oscillator loosely coupled to it.
3. A sufficient value of inductance L_C having a value X_2 equal to X_1 is then connected into the circuit as shown in Fig. 92.
4. The line is then checked for stationary waves, the absence of which indicates a condition of correct termination.

The mechanical properties of long concentric-tube transmission lines makes the measurement of current in the center conductor rather difficult. In some cases removable plugs are placed in the outside tube at various intervals along the line. These plugs, which, when inserted, make the outer tube airtight, permit connections from an antiresonant circuit across the line. Such an antiresonant circuit, when tuned to the fundamental frequency, presents a very high impedance to the line, when bridged across it, and therefore does not affect its characteristic impedance at the fundamental frequency. With about 10 watts flowing through the line, the galvanometer reading is an indication of the voltage at the points measured along the line.

73. Method Used in Measuring Antenna Characteristics. Refer to Fig. 98.

C_0 (usually about 0.0005 μf) is selected to provide sufficient series capacitance reactance to make the antenna capacitive over the frequency range measured. Then, with the antenna excited by the driver oscillator at the frequency indicated by the wavemeter and the switch at position 1, adjust C_1 and L_2 for resonance, as indicated by the maximum reading of G . R is then adjusted until G reading is the same as before. Then R is the antenna resistance.

For antenna reactance measurement, the circuit is first calibrated for stray capacitance in the shielded resistance box by resonating circuit (switch in position 2) first with box in the circuit and then entirely removed.

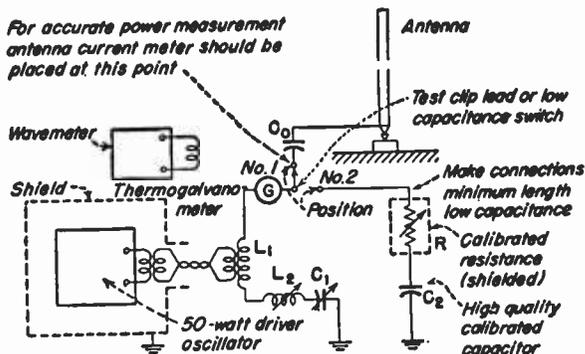


FIG. 98. Setup for measuring antenna characteristics.

Difference in reading of capacitor C_2 between the two conditions equals capacitance of box. This value should be added to each reading of C_2 , when circuits are resonated, which is done as above for resistance measurement. The antenna reactance X_2 is equal to the reactance of C_2 minus that of C_0 . When reactance of C_0 is greater than that of C_2 the antenna reactance is positive.

When it is found desirable to apply the matching circuits described above (Fig. 92) to balanced lines (open wire or double concentric types), the value of X_3 derived by the particular formula for cases 1 and 2 is halved and placed on each side of the circuit (see Fig. 99), while the value of X_4 is halved and placed on each side of the circuit for case

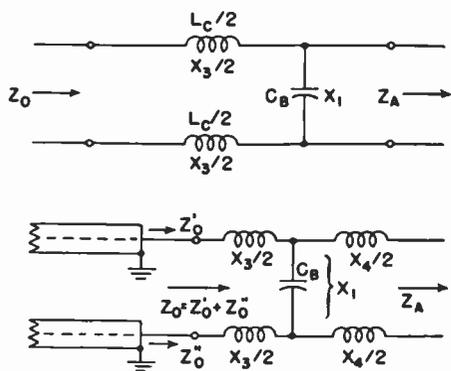


FIG. 99. Matching circuits for balanced transmission line.

3. Under these conditions the formulas given above apply to the respective cases mentioned. The systems become quite useful in matching a given balanced transmission line or r-f circuit into another having entirely different input impedance characteristics.

In the foregoing analysis of antenna matching circuits, they were considered as providing, for a given transmission line, a termination impedance equivalent to an ohmic resistance at the fundamental frequency. An analysis of the input impedance that such a line "looks into" at various harmonic frequencies discloses that it may assume an infinite number of different impedances containing resistive

and positive or negative reactance components, the values of which depend upon the termination circuit constants as well as those of the antenna. The values of antenna resistance and reactance may vary widely with frequency. For harmonic frequencies, stationary waves of current and voltage will form on the transmission line as

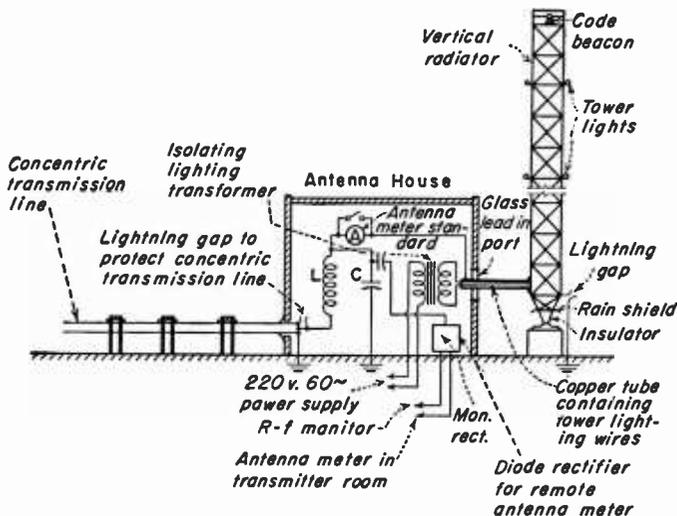


FIG. 100. Equipment in antenna house of modern broadcast transmitter.

well as in the antenna circuit, unless suitable harmonic filtering is provided either within the vacuum-tube transmitter or at the input to the transmission line.

The effectiveness of a given filter design for various harmonics depends upon its position in the line with respect to the positions of current and voltage antinodes of the harmonic frequencies along the line.

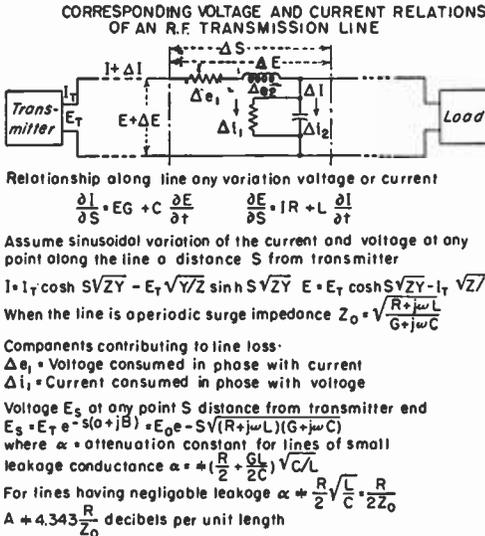
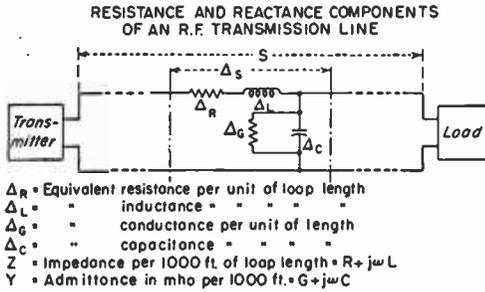


Fig. 101. Losses in transmission lines.

74. Loss in R-f Transmission Lines. From Fig. 101, it is evident that the most prominent factors contributing to power loss in open wire transmission lines are as follows:

1. Power loss due to conductor thermal resistance

$$R_T = \frac{0.1262}{d} \sqrt{\rho \mu f} \quad \text{ohms per cm length}$$

where $S \gg d$ (see Fig. 86)

- ρ = resistivity of conductors, microhm-cm
- μ = permeability of conductors
- f = frequency, Mc
- d = diameter of conductor, cm

2. Power radiated from balanced and unbalanced line currents
3. Power component of mutual inductance due to secondary currents induced in nearby conductors

4. Power loss due to leakage or conductance of the insulating medium or

$$G = \frac{\gamma A}{l} \quad \text{mhos per cm length}$$

5. Power loss due to dielectric hysteresis

For coaxial lines the major factors contributing to power loss in transmission lines are as follows:

1. Power loss due to conductor thermal resistance

$$R_T = 0.0631 \sqrt{\rho\mu f} \left(\frac{1}{d_1} + \frac{1}{d_2} \right) \text{ ohms per cm of line}$$

where d_1 = outside diameter of inner conductor, cm
 d_2 = inside diameter of outer conductor, cm
 ρ = resistivity of conductors, microhm-cm
 μ = permeability of conductors
 f = frequency, Mc

2. Power loss due to leakage conductance of insulating medium
3. Power loss due to dielectric hysteresis

By reference to the loss curves shown in Fig. 102, it is evident that the greater part of the power loss in both the open-wire and coaxial types when operating with neg-

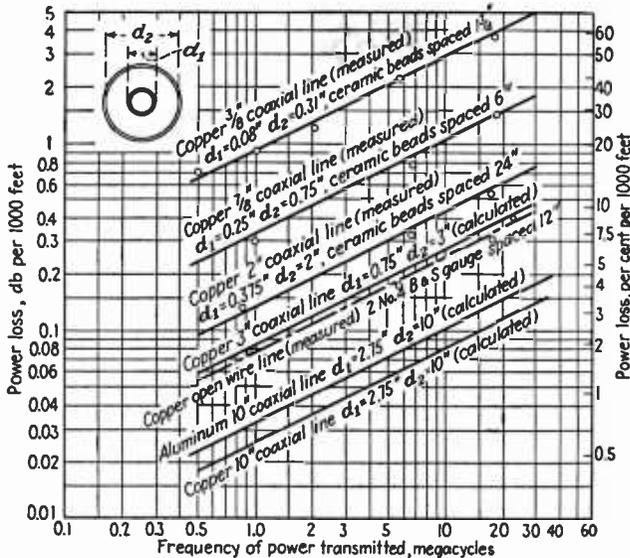


FIG. 102. Power loss in lines operating at 20°C with negligible reflection.

ligible reflection is due to the conductor thermal resistance. Owing to the low-loss insulation materials now available, the losses due to dielectric leakage and hysteresis can be reduced to a negligible quantity especially at standard broadcast frequencies.

The curves of Fig. 102 are the results of extensive r-f measurements which confirmed mathematical formulas given for calculation of losses in open-wire and coaxial lines of various standard sizes. Worthy of mention is the rather high efficiency of the open-wire line consisting of two No. 4 B&S gage solid copper conductors spaced 12 in. center to center. The losses in this line are less than those in a 3-in.-diameter copper coaxial line. The resistivity ρ of aluminum being greater than copper, the losses in an aluminum coaxial line are somewhat greater than those in a copper line of the same dimensions.

75. Broadcast-station Signal Coverage. The reception of satisfactory signals from a given broadcasting station by a particular listener at a given point depends upon the following: (1) the intensity of the signal radiated from the antenna system of the station as influenced by the radiated r-f carrier power, antenna directivity, and percentage of modulation on the carrier; (2) distance between the broadcasting-station antenna

and the point of reception and the attenuation characteristics of the intervening space or terrain; (3) intensity of objectionable interference at the receiving point; (4) fading as produced by the rays of direct and indirect signals; (5) the quality of the broadcast receiver and its ability to discriminate against local noise or interference and against adjacent channel interference and to convert the received r-f signals into sound without appreciable distortion or inherent receiver noise. The surrounding area about a given broadcasting station wherein satisfactory program signals can be received determines the service area of the station.

The service area of a standard broadcasting station of the 550- to 1,600-ke band consists essentially of two distinct regions. That region in close proximity to the station is served by the direct ray or ground wave called the *primary coverage area* of a broadcasting station, while the region at some distance from the station and served by virtue of indirect ray or sky-wave reflections is called the *secondary coverage area*. During daylight hours of broadcast transmission on frequencies of the standard broadcast band (between 550 and 1,600 ke), a broadcast listener is concerned with the primary coverage area signals of nearby stations for programs since there is very little sky-wave energy reflected during this period under normal conditions. The daylight service area of such a broadcasting station therefore consists almost entirely of that region served by the direct ray.

During the hours of twilight and darkness, the secondary coverage area of stations in the standard broadcast band becomes apparent. The secondary coverage area of a particular station begins at a considerable distance from a given station and is served by the predominant sky wave. The primary and secondary coverage areas of a standard broadcasting station are separated by a region known as the *fading area* of the station. In this area the signal intensities of the direct and indirect rays approach an equality with a result that violent fluctuations in signal intensities are apparent. The fading areas of stations are dependent upon a number of factors, such as frequency of transmission, antenna radiation characteristics, conductivity of intervening terrain, and time of day and season, and are independent of the transmitter powers of the stations. The fading area is normally in the form of a band about the broadcasting station normally contained within radii of between 20 and several hundred miles, depending upon the factors mentioned. The fading band may be as much as 50 miles in width.

Considering a standard broadcasting station radiating equally in all directions over surrounding terrain, and assuming equal ground attenuation, the service area would consist of a primary coverage area or circular area near the station and served by a steady ground-wave signal. Outside of this would exist the fading area consisting of a ring about the primary area. Beyond the fading ring the secondary coverage area would exist. Inasmuch as broadcast reception is rather uncertain in the fading region and in the secondary coverage area, the real value of a given station is dependent normally upon its primary coverage area.

The primary service area of a particular station can be most accurately determined by means of a field-intensity survey. A survey¹ of this kind is accomplished through the use of mobile field-intensity measuring equipment. This consists essentially of a field-intensity meter or carefully shielded receiver equipped with an indicating meter at its output terminals to read carrier-signal intensity as induced in the loop antenna. The field-intensity meter together with the loop antenna are carefully calibrated in their position in the measuring car to give accurate readings in microvolts per meter over a wide range of carrier-signal intensity.

76. Field-intensity Measurements. The procedure of making a field-intensity survey of a standard broadcasting station consists usually of making frequent measurements at satisfactory positions (in free space) along radials progressing to and from the station. Eight or more radials at equal angular spacing are generally made about a point established on the field survey map by the broadcast station antenna system

¹ Kirby, S. S., and K. A. Norton, *Field Intensity Measurements*, *Bur. Standards J. Research*, April, 1932.

and extending to a signal intensity of 500 mv or beyond. Each radial is then plotted on loglog coordinate graph paper and a smooth curve drawn through these points to show directly the signal intensity along one ordinate, with distance along the other. Later the values required are transferred to a map in the form of signal contour lines representing positions about the station where field intensities of 100, 50, 10, 2, and 0.5 mv per m exist. The contour map for reference purposes also contains information such as (1) station call letters, (2) frequency, (3) antenna power and its directivity and other characteristics, (4) scale of map, and (5) date.

Since fading occurs after sunset, these measurements are an indication of satisfactory daytime coverage only from the particular station. As recommended by reports of the IRE,¹ FCC,² and the National Association of Broadcasters, values of standard broadcast field intensity considered necessary for reliable broadcast service are given for three areas as follows: (1) a business city area where a field intensity of from 10 to 25 mv per m is required to override high interfering electrical noise and overshadowing effects of large buildings, (2) a residential district of a city where a field intensity of 2 to 5 mv per m is required, (3) a rural area where 0.1 to 0.5 mv per m signal intensity is sufficient. In addition it is stated that for fair service a signal intensity of one-half the above values is needed and for poor service one-fourth of these values. These figures are based upon the average signal intensity necessary to override the noise levels of these districts. In large cities, where large, tall buildings are numerous, a free-space field intensity of as much as 50 mv per m over the city may be necessary to provide a signal intensity at a particular receiving antenna between buildings of one-fifth of that amount.

Since the primary service area of a standard broadcasting station includes nighttime reception as well as daytime, fading measurements are necessarily a part of the field-intensity survey in determination of this area. Fading measurements are made with the same field-intensity measuring equipment used for the survey except that the field-intensity meter is equipped with a recording milliammeter (usually of 0 to 5 ma range) attached to the output of the field-intensity meter. A d-c amplifier sometimes is necessary to secure sufficient signal level to actuate the recording meter from the field-intensity measuring set. The equipment is set up for periods of time at a given distance and location from the station, and fluctuations in carrier-signal intensity are noted on the continuously moving recording chart. Amplitude fluctuations as recorded on the chart indicate the amount of fading. Fading measurements for considerable periods of time and over a wide area are necessary to determine the fading region about a given station and to evaluate the secondary coverage area about the station, particularly those designated as class I stations.

Field-intensity measurements for v-h-f broadcasting stations are made in a manner similar to that for standard broadcasting with the exception that a different type of field-intensity measuring equipment is required and that the procedure for making fading measurements is not generally required.

77. Calculations of Station Coverage. A mathematical investigation of the attenuation of radio waves propagating over plane earth has led to mathematical expressions which follow very nearly the characteristics of waves as indicated by actual measurements. A simplified form of this expression requires the following information for a solution: (1) the frequency of the transmitted wave, (2) the distance from the station, (3) conductivity of the soil in electromagnetic units, and (4) the inductivity of the soil in electrostatic units. Since the inductivity can be generally assumed to be 14 to 15 esu, then, with a measured value of conductivity σ , the field intensity at a given distance from a station may be calculated. With further assumptions concerning the irregularities in general characteristics of the terrain about the station, it is possible to calculate the contours. The value of σ (the soil conductivity) is usually secured from a measured radial or taken from available field-intensity measurements

¹ Report of Committee on Radio Propagation Data, *Proc. IRE*, 21 (10), October, 1933.

² Fifth annual report to the Congress of the United States, by Federal Communications Commission, gives tabulated values of field strength.

of some other station in the vicinity. The FCC has published charts showing soil conductivity over the United States. Provided measured values are not available, these may be used.

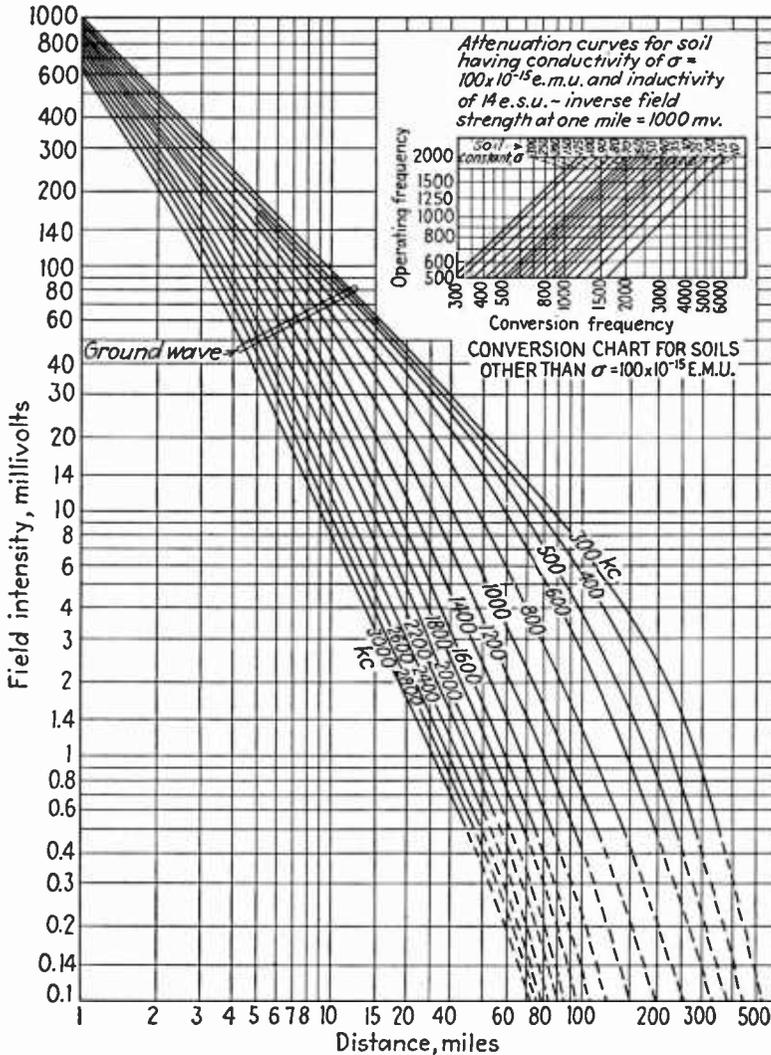


FIG. 103. Ground-wave field-intensity curves.

For convenience the chart shown in Fig. 103 is given. It may be used to calculate signal attenuation of standard broadcast frequencies. The attenuation curves shown are derived from a simplified form of Sommerfeld's attenuation formula.¹ With a

¹ Sommerfeld, Arnold, Ausbreitung der Wellen in der drahtlosen Telegraphie Einfluss der Bodenbeschaffenheit, und gerichtete und ungerichtete Wellenzüge, *Jahrb. drahtlosen Tele. Tele.*, 4, December, 1910. Rolf, Numerical Discussion of Sommerfeld's Attenuation Formula, *Proc. IRE*, 18 (3), March, 1930. Eckerley, P. P., The Calculation of the Service Area of Broadcast Stations, *Proc. IRE*, 18 (7),

single set of Sommerfeld curves to cover all the standard broadcast frequencies and soil conductivities, the conductivity of a given soil can be rather easily computed from the attenuation of a particular signal. This is accomplished by first converting a given radial to an inverse field strength of 1,000 mv at 1 mile and then determining the frequency of the ground-wave curve with which it coincides, *i.e.*, the conversion frequency. The conductivity is secured from the soil-constant curve passing through the intersection of the operating and conversion frequencies on the conversion chart in the upper right corner of the figure. The conversion chart has been prepared from the following relationships:

$$f_1 = f \sqrt{\frac{\delta}{\delta_1}}$$

where f = operating frequency

δ = standard conductivity of chart (100×10^{-16} emu)

δ_1 = actual soil conductivity

f_1 = conversion frequency

For example, assume a station operating on an assigned frequency of 660 kc where the field strength radial, as plotted from measurements, follows the 1,500-kc curve. Then, from the soil conversion chart, the conductivity is very nearly 20×10^{-16} emu. On the other hand, if the soil conductivity is known, the signal attenuation can be determined from the conversion chart and the attenuation curves of various frequencies. Since these curves are based on a field strength of 1,000 mv at 1 mile, the actual signal at a given distance from a station is, of course, derived from the ratio of the actual signal intensity in millivolts at 1 mile from the particular station divided by 1,000. At considerable distances from the transmitter these curves are subject to corrections for the effects of curvature of the earth.

The curves in Fig. 104 refer to sky-wave intensities under various conditions of propagation. These are most useful in the determination of the fading regions about a particular station and are plotted to give intensities of reflected sky-wave intensity for different antenna electrical heights based on a signal intensity of 1,000 mv along the ground at 1 mile from a given antenna. In this case the electrical height of the antenna in degrees equals $3.85 \times 10^{-4} Hf$, where H is the physical height of the antenna in feet and f is the operating frequency in kilocycles. This is based on a velocity of propagation equivalent to 0.95 that of light.

Inasmuch as the attenuation curves of ground-wave intensity (Fig. 103) are also based on 1,000 mv at 1 mile, then the particular distance from a given antenna where the sky-wave intensity, shown on curves of Fig. 104, equals the ground-wave signal intensity of Fig. 103 is the distance from the antenna where one would expect to observe greatest fading or is an estimate of the center of the fading band. Owing to the height of the Heaviside layer being other than 100 km and reflection being less than unity, on which these curves are based, calculated distances given by these curves are approximate.¹ Measurements are required for more exact determination of the fading region.

The service rendered by a standard broadcast station depends also on interference caused by other stations on the same and near-by channels. This interference is greatly increased at night because signals from undesired distant stations are reflected by the Heaviside layer and may be received with varying intensities within the service area of a desired station. Following extensive survey work covering nighttime signal

July, 1930. Eckersley, T. L., Direct Ray Broadcast Transmission, *Proc. IRE*, 20 (10), October, 1932. Norton, K. H., Propagation of Radio Waves over a Plane Earth, *Nature*, June 8, 1935; Propagation of Radio Waves over the Surface of the Earth and in the Upper Atmosphere, Part I, *Proc. IRE*, 24, October, 1936; Part II, *Proc. IRE*, 25, September, 1937. Fitch, W. A., The Sommerfeld Formula, *Electronics*, 9 (9), September, 1936. For FCC curves of standard broadcasting-station propagation, refer to curves of Appendix 1, "Standards of Good Engineering Practice concerning Standard Broadcast Stations," U.S. Government's Printing Office, Washington.

¹ See Amendments to Standards of Good Engineering Practice concerning Standard Broadcasting Stations, *FCC Public Notice*, 4072, Jan. 17, 1947.

propagation over the period February to May, 1936, the FCC issued a report¹ wherein a great amount of information concerning sky-wave propagation is given. In Fig. 105 are illustrated curves representing the average sky-wave field intensity (second hour

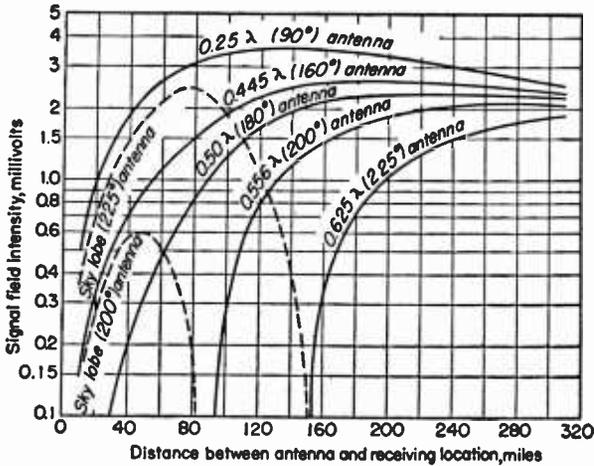


FIG. 104. Sky-wave propagation curves for antennas of various heights. Layer height = 100 km; reflection coefficient = 1.

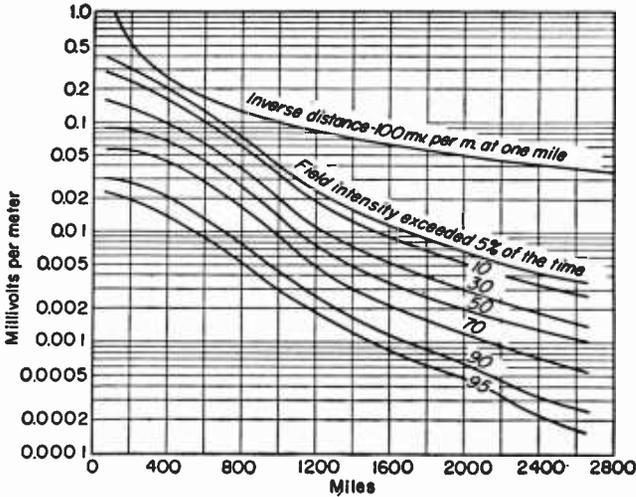


FIG. 105. Average sky-wave field intensity, 640 to 1190 kc.

after sunset) at the recording station. An interfering or undesired signal existing for 10 per cent of the time has been standardized as an interfering signal. Thus, with the antenna sky-radiation characteristics of a given station known,² it becomes possible to estimate the amount of undesirable interference it is liable to cause to another dis-

¹ FCC Report 18108, September, 1936.

² Standards of Good Engineering Practice concerning Standard Broadcast Stations, FCC Report 41831, June 29, 1940.

tant station.¹ In the determination of interference problems, the FCC has issued as an amendment, dated Aug. 8, 1949, to the Standards of Good Engineering Practice Table 4 to be used for determining the minimum ratio of the field intensity of a desired to an undesired signal for interference-free service.

Table 4. Interference Ratios

Frequency separation of desired to undesired signals, kc	Desired ground wave to		Desired 50 per cent sky wave to undesired 10 per cent sky wave
	Undesired ground wave	Undesired 10 per cent sky wave	
0	20:1	20:1	20:1
10	1:1	1:5	
20	1:30		

When the desired ground-wave signal is interfered with by two or more sky-wave signals on the same frequency, the r-s-s-value of the latter is used.

Furthermore the rules state as follows:

... accordingly, no station will be licensed for operation with less than 40 kc separation from another station if the 25 mv/m ground-wave contours of the two stations overlap. Frequency separation of 20 kc and 10 kc are considered inappropriate for stations with the same urban coverage and therefore no station will be licensed with less than 30 kc frequency separation if the area enclosed by the 25 mv/m ground-wave contour of either one overlaps the area enclosed by the 2 mv/m ground-wave contour of the other.

The FCC has classified standard broadcast stations with respect to protected service contours and permissible interference signals in accordance with Table 5.

Table 5. Protected Service Contours and Permissible Interference Signals for Broadcast Stations

Class of station	Class of channel used	Permissible power, kw	Signal intensity contour of area protected from objectionable interference, $\mu\text{v}/\text{m}$		Permissible interfering signal on same channel, $\mu\text{v}/\text{m}$	
			Day*	Night†	Day*	Night†
I-A	Clear	50	SC 100 AC 500	Not duplicated	5	Not duplicated
IB	Clear	10-50	SC 100 AC 500	500	5	25
II	Clear	0.25-50	500	(50% sky wave) 2,500†	25	125‡
III-A	Regional	1-5	500	(ground wave) 2,500	25	125
III-B	Regional	0.5-1 per night and 5 per day	500	(ground wave) 4,000	25	200
IV	Local	0.1-0.25	500	(ground wave) 4,000	25	200

* Ground wave. SC = same channel; AC = adjacent channel.

† Sky-wave field intensity for 10 per cent or more of the time.

‡ These values are with respect to interference from all stations except class I-B, which stations may cause interference to a field-intensity contour of higher value. Class IX stations may also be assigned to regional channels according to Sec. 3.29.

77. V-h-f broadcast-station coverage concerns the stations licensed by the FCC primarily for the transmission of radiotelephone emissions in the v-h-f broadcast band for reception by the general public. The v-h-f broadcast band contains the band of frequencies extending from 88 to 108 Mc. In accordance with Standards of Good Engineering Practice concerning FM Broadcast Stations of the FCC, the stations in this band must use a system of modulation of the radio signal in which the frequency of the carrier wave is varied with the program signal; this being commonly termed *frequency modulation*, or f.m. The assigned operating frequency or "center frequency" is that of the r-f carrier without modulation. It must be maintained within 2,000 cycles of the assigned center frequency assigned. Channels for v-h-f broadcast sta-

¹ Modified by FCC Public Notice 4072, Jan. 17, 1947, with curves.

tions begin at 88.1 Mc and continue in successive steps of 200 kc to and including the assigned frequency of 107.9 Mc.

According to the FCC rules, v-h-f broadcast stations shall be licensed on the basis of an area in square miles within the service area. The contour bounding the service area and the radii of same are determined in accordance with the FCC standards. On this basis, a v-h-f broadcast station has a single service; that corresponding to the primary service of a standard broadcast station. Secondary, sky-wave or intermittent, service is not recognized in v-h-f broadcast coverage.

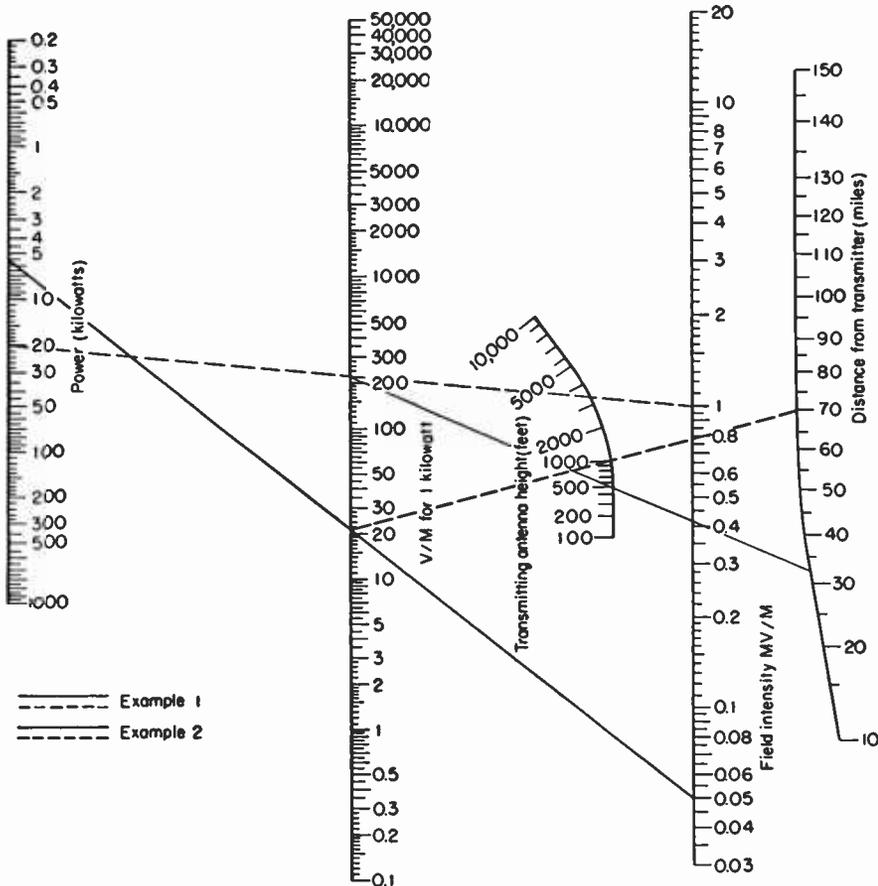


FIG. 106. Signal-range chart for v-h-f broadcast stations.

H-f broadcast transmitters are normally located as near to the center of the proposed service area as possible. A high elevation of the transmitting antenna is necessary to reduce the shadowing effects upon propagation due to hills, buildings, and other obstructions in the city area. The position of the transmitter site is also dependent upon the purpose of the station, *i.e.*, whether it is intended to serve a small city, a metropolitan area, or a large region. A suitable transmitter site may be made available by the use of a directive antenna. Where a directive antenna is used, a centrally located station site may not be a desirable one. As one may understand by studying

the chart in Fig. 106, the transmitter antenna height above the average elevation of the service area is a consideration of greatest importance to secure optimum coverage with a h-f broadcast station.

In the consideration of objectionable interference from other stations on the same and adjacent channels, the FCC rules require that the proposed station shall not have interference to such an extent that its service may be reduced to an unsatisfactory amount. For this reason, objectionable interference is considered to exist when the interfering signal exceeds that given by Table 6. In this table the desired signal is median field, and the undesired signal is the tropospheric signal intensity exceeded for 1 per cent of the time.

Table 6. Allowable Signal Ratios (FCC, Aug. 10, 1948)

Channel Separation	Ratio of Desired to Undesired Signals
Same channel.....	10:1 median field intensity
Adjacent channel (200 kc removed).....	2:1 median field intensity

Objectionable interference is not considered to exist when the channel separation is 400 kc or greater. Therefore f-m broadcast stations in the same city or area may be assigned channels 400 kc apart.

The service area of f-m broadcasting stations is considered to be only that served by the ground wave. The extent of the service is determined by the point at which the ground wave is no longer of sufficient intensity to provide satisfactory service. The standard of field intensity necessary for satisfactory service is given as follows:

Table 7. Service

Area	Median Field Intensity mv/m
City areas near factories, car lines, or busy streets.....	1
Rural areas away from highways.....	0.050

The FCC standard states

... a median field intensity of 3,000 to 5,000 μv per m should be placed over the principal city to be served and a median field intensity of 1,000 μv per m should be placed over the business district of cities of 10,000 or greater within the metropolitan district served.

To determine the approximate distance to a particular contour of v-h-f f-m station the chart in Fig. 106 may be used. The results obtainable are based on a signal intensity at a receiving antenna with an elevation of 30 ft. The distance to the 50 μv per m contour about a given station is dependent upon values of the transmitting antenna height, the antenna power, and the antenna gain. This chart is prepared for a carrier frequency of 98 Mc in the center of the f-m band and may be used for all f-m broadcast channels in the 88- to 108-Mc band with sufficient accuracy since little change results over this frequency range.

Examples: Two examples are offered to illustrate the use of the chart. Consider, (1) a northeastern metropolitan station and (2) a transmitter site 1,000 ft high with a 50- μv signal required at a distance of 70 miles from the transmitting antenna.

1. The radiated power of the northeastern station is assumed to be 20 kw at an antenna height of 500 ft. The problem is to predict the range in miles to the 1-mv contour. In Fig. 106, the dotted line between 20 kw and the 1 mv per m field intensity is drawn. This line intersects the microvolt plot for 1 kw at 224 μv . A solid line then drawn between this point of intersection and the 500-ft transmitting antenna height point shows a distance of 32 miles.

2. Assume that there is required a 50- μv signal 70 miles away. Draw, between the 70-mile point and the point indicating 1,000 ft, a line which indicates that a radiated power of 1 kw from a station using an antenna 1,000 ft high would provide 21.5 μv at a distance of 70 miles. A solid line drawn between this point, the 0.050-mv point, and the power scale indicates that a transmitter power output of 5.5 kw is required to give a field intensity of 50 μv (0.050 mv) at a distance of 70 miles from the antenna radiating at an elevation of 1,000 ft above average ground at the transmitting station.

CHAPTER 19 TELEVISION

BY DONALD G. FINK¹

1. Definition. Television is the electrical transmission of transient visual images. Cathode-ray television makes use of electron beams or electron images in the camera tube (pickup device) and in the picture tube (reproducing device). A television system possesses facilities for transmitting sound synchronously with visual images.

2. Elements of a Television System. The elements of a typical television system are shown in Fig. 1. The sound system consists of an f-m transmitter and receiver operating with 25-kc maximum deviation and is separate from the picture system,

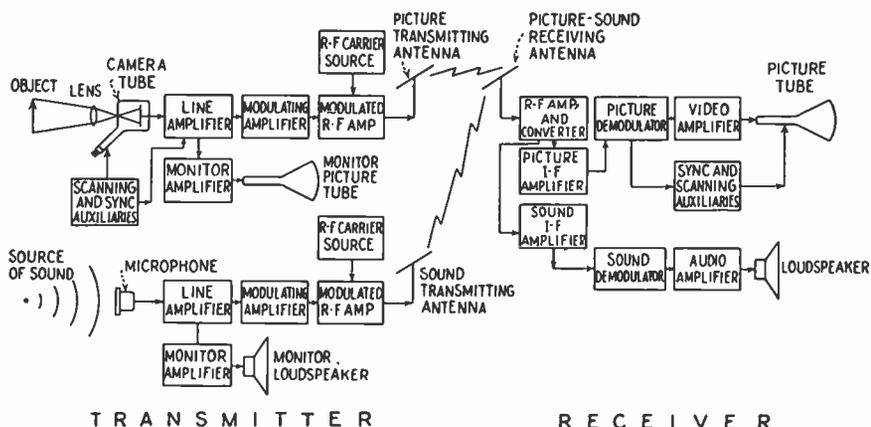


FIG. 1. Elements of television system.

except that common antennas may be employed at the transmitter and receiver and a common r-f amplifier and first detector may be used in the receiver. The picture transmitter includes the camera and synchronization circuits (video signal generator), video amplifiers, u-h-f carrier source and r-f amplifiers, the modulator, a filter for suppressing part of one of the side-band regions in the carrier output, and the radiator. The picture receiver consists of r-f amplifier, antenna circuits, first detector and i-f amplifiers (the latter two in superheterodyne receivers), a second detector, one or more video amplifiers, picture tube, synchronizing signal-separator circuits, scanning generators, and power supplies.

SCANNING AND IMAGE ANALYSIS

3. Linear Scanning. The method of analyzing and synthesizing visual images employed in modern television systems is known as *linear scanning*. As applied to the transmission of images, linear scanning involves the exploration of the image to be transmitted by an elemental spot of small area, known as the *scanning agent*, which

¹ Editor, *Electronics*.

traverses the area of the image in a series of horizontal lines, moving over every point in the image at constant speed and discovering the degree of brightness at each point in succession. The camera tube, which includes the scanning agent, generates a succession of electrical impulses which correspond with the successive values of brightness discovered by the scanning agent.

At the receiver the scanning process involves setting up an elemental luminous spot of small area which moves synchronously with the scanning agent in the camera tube. The brightness of this luminous spot is controlled by the electrical impulses transmitted from the camera tube to the receiver. The values of brightness present in the original image are thereby reproduced in their proper positions. The scanning process must be rapid enough so that all the elements of the received image are perceived simultaneously by the eye. This requirement is met if the scanning of the image is completed within the duration of persistence of vision, so that the first element of brightness persists in the eye during the production of all the succeeding elements in the image.

4. Aspect Ratio. The ratio of width (w) to the height (h) of the rectangle actively employed in reproducing the image is known as the *aspect ratio*. In accordance with the standard adopted for motion pictures, in the United States this ratio is given the value

$$\frac{w}{h} = \frac{4}{3} \tag{1}^*$$

5. Total Number of Lines per Frame. The total number of lines over which the scanning agent passes from the beginning of one complete image to the beginning of the next is known as the total number of lines per frame, n .

The number of lines determines the degree of detail which may be accommodated in the reproduced picture, in the vertical dimension. Hence this number sets an upper limit to the amount of pictorial detail which may be accommodated in that dimension. The number in modern systems is set usually between 400 and 700 lines. According to the FCC standards, n has the value

$$n = 525 \tag{2}^*$$

The reason for the exact number 525 (see Secs. 6 and 31) is that it is an odd number composed of simple odd factors

$$525 = 3 \times 5 \times 5 \times 7$$

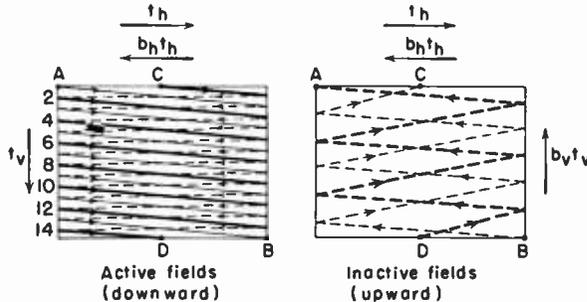


FIG. 2. Scanning pattern for two-to-one odd-line interlaced scanning.

6. Interlaced Scanning. To reduce flicker in the reproduced image, a scanning technique known as *interlacing* is customarily employed, whereby the image is scanned in two groups of lines. The scanning motion in "two-field odd-line" interlaced scanning (the method now universally adopted) is shown in Fig. 2. The scanning agent

* Those relationships marked with an asterisk (*) are the official standards prescribed by the FCC for commercial television broadcasting.

traverses the area in two series of lines, alternately, passing downward (at left in Fig. 2) from point A to B in the light solid line, following the back-and-forth motions shown by the arrows. The scanning spot then moves upward from point B to C (at right), thence downward again from point C to D on the heavy line, finally upward again from point D to the starting point A , where the motion repeats itself. The scanning agent is active in discovering or reproducing the picture information while traveling over the lines shown solid and is inactive while traveling over the lines shown dashed. The total number of back-and-forth motions made in traversing both series of lines is n . The total number of active lines (shown solid) is n_a . The inactive lines ($n - n_a$) are those made by the scanning agent in traveling from the bottom to the top of the picture area (the motions shown at the right). The number of active lines, n_a , for a 525-line image is approximately 490. The exact value of n_a is given by

$$n_a = \frac{n}{1 + b_v} \quad (3)$$

where b_v , the *vertical retrace ratio*, is the ratio between the upward and the downward scanning time, as defined in Sec. 7.

7. Scanning Velocities and Retrace Ratios. The scanning agent is made to traverse the picture area in the interlaced pattern (Fig. 2) by imparting to it horizontal and vertical motions. The spot is displaced horizontally from left to right during a time interval t_h sec, and simultaneously it is displaced vertically downward over a longer time t_v sec. The two motions cause the spot to move slightly downward and to the right until it reaches the right-hand edge of the area. Then the spot is reversed rapidly and is moved to the left in a fraction of the time occupied in moving from left to right. This fraction is the horizontal retrace ratio, b_h . Thus the complete back-and-forth horizontal motion occupies a time of $t_h(1 + b_h)$ sec. The downward motion persists during the succession of back-and-forth motions until the spot reaches the bottom of the area at the end of t_v sec. Thereupon the downward motion is reversed, and the spot is moved upward in a fraction, b_v , of the time consumed in the downward motion, until it reaches the top of the pattern. During the upward motion, several back-and-forth motions are executed, since the horizontal motions are maintained.

The *horizontal retrace ratio* b_h is the ratio of the backward (to the left) scanning time to the forward (to the right) scanning time. Practical values of b_h range from about 0.10 upward to 0.25. The FCC standards set an upper limit to this ratio of 0.15. The *vertical retrace ratio*, b_v , between the upward and the downward times, ranges from about 0.05 to 0.10. The FCC standards set an upper limit to this ratio of 0.08.

8. Interlaced Fields. One set of the two sets of lines in the interlaced pattern is known as an *interlaced field*. Since the total number of lines in the complete frame is an odd number (525), the number of lines per field is a whole number plus one-half ($262\frac{1}{2}$). This accounts for the fact that at the end of the first field (Fig. 2) the spot, at point C , has formed but one-half the horizontal motion. The half-line displacement causes the lines in the second field to be displaced vertically from those in the first field by the width of one line. Consequently the lines of one field fall directly between the lines of the preceding field. If the scanning motion is not accurately timed and if the amplitudes of the vertical and horizontal motions are not constant, this interlaced relationship is not maintained, and the lines of one field tend to overlap the lines of the preceding field. This defect is known as *pairing* of the fields. Its effect is to reduce the detail of the reproduced picture in the vertical dimension.

9. Vertical Resolution. The vertical resolution r_v of the scanning pattern is measured by the number of pictorial details or *picture elements* which may be accommodated in the vertical height of the picture area. Each active scanning line is capable of reproducing one such picture element in the vertical direction but, since the picture elements in the image to be transmitted may not fall directly on the scanning lines, the actual number of picture elements which may be accommodated vertically is less than the number of active scanning lines. The vertical resolution r_v is accordingly the

number of active scanning lines multiplied by a factor less than 1, the utilization ratio k . The vertical resolution is accordingly

$$r_v = kn_a \quad \text{elements per picture height} \quad (4)$$

Practical values of the utilization ratio, depending on the perfection of interlacing, range from about 0.6 to 0.9. With $n_a = 490$, r_v accordingly varies from 300 to 440 elements per picture height. The value $r_v = 350$ is commonly attained in properly operated equipment.

10. Horizontal Resolution. The horizontal resolution r_h of the scanning pattern is

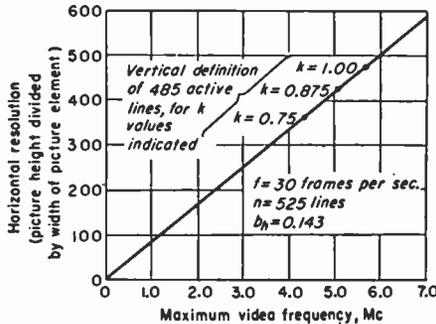


FIG. 3. Relationship between horizontal resolution and maximum video frequency (441-line image).

measured by the number of picture elements which may be accommodated in the horizontal direction, measured in a width equal to the picture height. The picture height is used as the basis to make the horizontal resolution directly comparable with the vertical resolution. The total number of picture elements accommodated in the picture width is the horizontal resolution multiplied by the aspect ratio.

The value of r_h does not depend on the dimensions of the scanning pattern but rather on the electrical performance of the television system in reproducing rapid changes of voltage whereby the reproducing scanning agent is changed in brilliance as it moves across each

line. In terms of the maximum frequency f_{max} in the video range (see Sec. 16), the horizontal resolution is approximately

$$r_h = 84f_{max} \quad \text{elements per picture height} \quad (5)$$

where f_{max} is expressed in megacycles. This expression assumes transmission at a rate of 30 frames per second and 525 lines. Horizontal resolution of 340 lines ($f_{max} = 4.0$ Mc) is attainable in properly operated equipment.

11. Resolution Ratio. The ratio of the horizontal resolution to the vertical resolution is the resolution ratio m :

$$m = \frac{r_h}{r_v} = \frac{84f_{max}}{kn_a} \quad (6)$$

Unity resolution ratio (equal resolution in vertical and horizontal directions) is not essential for good reproduction; *i.e.*, the resolution in one direction may exceed that in the other without wasting the detail in the direction of higher resolution. In present practice the resolution ratio approaches 0.95, depending on the maximum frequency in the video range. For values of $f_{max} = 4$ Mc, $n_a = 485$ lines, and $k = 0.75$, the resolution ratio is 0.925.

12. Total Number of Reproducible Picture Elements. A significant figure of merit of the television system is the total number, N , picture elements which may be accommodated in the picture area, *i.e.*, the product of the number of elements vertically r_v , times the number horizontally $(w/h)r_h$:

$$\begin{aligned} N &= \left(\frac{w}{h}\right) r_h r_v = \left(\frac{w}{h}\right) (84f_{max}) (kn_a) \\ &= \left(\frac{w}{h}\right) mk^2 n_a^2 \end{aligned} \quad (7)$$

For values of $(w/h) = \frac{4}{3}$, $f_{max} = 4.0$ Mc, $n_a = 485$ lines, and $k = 0.75$ ($m = 0.925$),

the total number is $N = 165,000$ picture elements. Performance above 200,000 picture elements is exceptional in the present state of the art.

13. Viewing-distance Relationships. The desirable viewing distance of a television image depends on the resolution available. If we assume a visual acuity of 1 min of

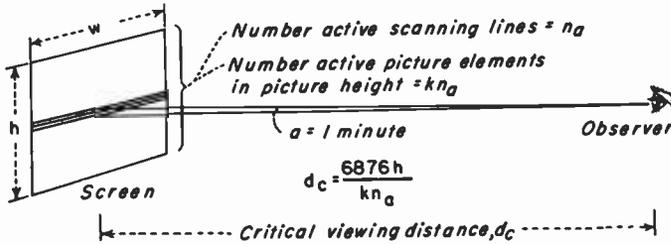


FIG. 4. Critical viewing distance in terms of the dimensions of the scanning pattern.

arc (typical of most normal eyes), two stationary dark picture elements separated by a bright element (Fig. 4) may be barely resolved by the eye at a distance d_c .

$$d_c = \frac{6,876h}{kn_a} \tag{8}$$

and the corresponding ratio of critical viewing distance to picture height is

$$\frac{d_c}{h} = \frac{6,876}{kn_a} \tag{9}$$

For a vertical resolution $r_v = kn_a$ of 350 elements per picture height, the foregoing ratio is about 20 times. This is the maximum viewing distance (20 times the picture height), beyond which the eye is unable to resolve the detail actually present in a stationary image. In practice, since the picture elements are not sharply defined and since the act of following motion reduces visual acuity, full detail can rarely be resolved at distances greater than eight times the picture height.

The minimum viewing distance is determined by the tolerance of the viewer toward the structure of the picture, which becomes increasingly evident as the viewing distance is decreased. Viewing distances shorter than three times the picture height are seldom considered satisfactory. A ratio of 5:1 seems to be typical of viewer habits. Figure 5 shows Eq. (9) for various numbers of scanning lines.

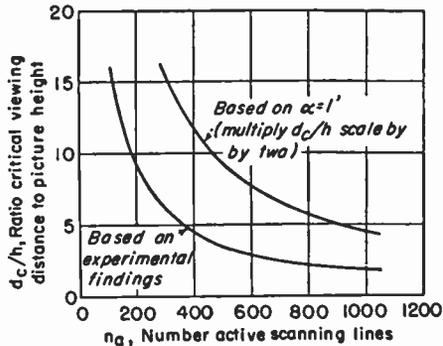


FIG. 5. Viewing distance vs. number of scanning lines, in terms of picture height.

14. Frame-repetition Rate. The rate at which the frames are repeated (frame-repetition rate) depends (1) upon the duration of the persistence of vision of the eye and (2) upon the necessity of reproducing motion in the image in a smooth manner. In motion pictures the standard rate is 24 frames per second, with each frame projected twice, making 48 projection intervals per second. Similar values serve for television. However, since the power-supply frequency for most areas in this country is 60 cps, it is desirable to use a frame-repetition rate f which is a submultiple of the power frequency, e.g., 30 per second (field repetition rate f' of 60 per second). The FCC

standards for these items are

$$f = 30 \text{ frames per second} \quad (10)^*$$

$$f' = 60 \text{ fields per second} \quad (11)^*$$

15. Rate of Scanning Picture Elements. The maximum rate of scanning picture elements along each line equals the number of elements in the line divided by the time during which the line is scanned. These quantities in turn depend on the horizontal resolution (Sec. 10) and on the number of lines per frame (Sec. 5) and the rate of frame repetition (Sec. 14). The general expression for the maximum rate of scanning picture elements R is

$$R = \frac{w}{h} m f k n^2 \frac{(1 + b_h)}{(1 + b_v)} \quad \text{elements per second} \quad (12)$$

where the quantities have been defined in the preceding sections. For aspect ratio $w/h = \frac{4}{3}$, resolution ratio $m = 0.925$, frame-repetition rate $f = 30$ per second, utilization ratio $k = 0.75$, number of lines per frame $n = 525$, horizontal retrace ratio $b_h = 0.20$, and vertical retrace ratio $b_v = 0.08$, the rate of scanning picture elements is approximately $R = 8,500,000$ elements per second, which is approximately the upper limit of performance of present-day equipment.

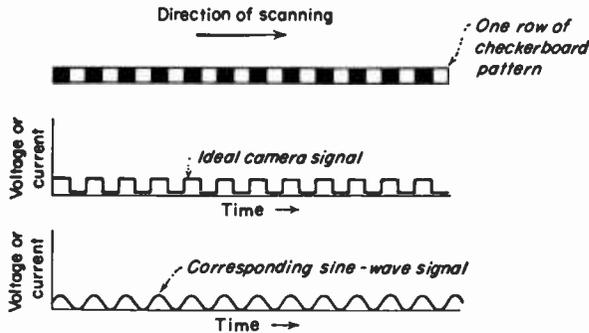


FIG. 6. Wave forms resulting from scanning a checkerboard image. The ideal square wave becomes a sine wave when only the fundamental frequency is transmitted.

16. Maximum Frequency in Video Range. The maximum video frequency generated by the television camera is directly proportional to the rate at which the picture elements are scanned along each line. In deducing a relationship between the scanning rate R (Sec. 15) and the maximum video frequency (v.f.), it is customary to assume that the picture elements are arranged as alternate black and white squares along the scanning line. An ideal scanning agent, scanning such a line, will produce a square wave, as shown in Fig. 6. The upper portion of each square wave represents a black element, the lower portion an adjacent white element. Hence there are two elements per cycle of the wave. The fundamental frequency of the square wave is accordingly one-half as great as the rate of scanning picture elements. The maximum v.f. is then derived from Eq. (12), as

$$f_{\max} = \frac{(w/h) m f k n^2}{2} \cdot \frac{(1 + b_h)}{(1 + b_v)} \quad (13)$$

For the conditions cited in Sec. 15, f_{\max} is 4.25 Mc. Table 1 gives other typical values. It should be noted that this frequency is the fundamental of the square wave. The reproducing equipment cannot reproduce a square wave of this frequency. Instead a sine-wave distribution of light is reproduced. This sine wave (Fig. 6) establishes the basic structure of the reproduced image.

17. Scanning Wave Forms. The deflecting forces necessary to produce the linear scanning motions shown in Fig. 2 are *saw-tooth waves*, as shown plotted against time

Table 1. Maximum Video Frequencies for Different Scanning Patterns

No. of scanning lines, n	No. of frames per second, f	Max v.f. for equal vertical and horizontal resolution ($m = 1.00$), cps	Max v.f. for horizontal resolution = $0.9 \times$ vertical resolution ($m = 0.9$), cps
20	16	3,360	3,020
60	16	30,200	27,250
120	24	81,500	163,000
180	24	410,000	369,000
240	24	727,000	653,000
343 ($7 \times 7 \times 7$)	30	1,860,000	1,670,000
441 ($3 \times 3 \times 7 \times 7$)	30	3,060,000	2,800,000
525 ($5 \times 5 \times 7 \times 3$)	30	4,350,000	3,920,000
1,029 ($3 \times 7 \times 7 \times 7$)	30	16,650,000	14,800,000

NOTE: Calculation based on $w/h = \frac{4}{3}$, $b_h = 0.143$, $b_v = 0.083$, $k = 0.75$.

in Fig. 7. The forward motion in the horizontal direction is produced by a deflecting force linear with time, and the retrace motion by a force which need not be linear but which must have a rate of change high compared with that of the forward force. The ratio of the slopes is equal to the inverse of the horizontal retrace ratio b_h (Sec. 7). The same conditions apply to the deflecting force in the vertical direction, and the ratio of the retrace slope to the forward slope of this wave equals the inverse of the vertical retrace ratio, b_v .

The scanning wave forms have fundamental frequencies determined by the number of fields per second and by the number of lines per second. In the vertical direction the scanning force must repeat itself at the field repetition rate, $f' = 60$ cps. In the horizontal direction the deflecting force must repeat itself at the line-scanning frequency (525 lines per frame, 30 frames per second), which is the product

$$nf = 525 \times 30 = 15,750 \text{ cps}$$

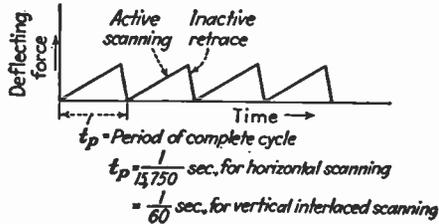


FIG. 7. Saw-tooth waves of deflecting force used to produce the scanning pattern (525-line image).

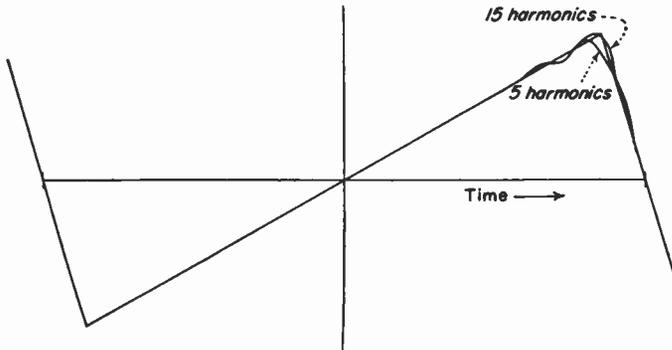


FIG. 8. Ideal saw-tooth wave and approximations resulting from inclusion of 5 and 15 harmonics.

These values of scanning frequency are the fundamentals of the saw-tooth wave. From 5 to 20 harmonics should be present if the wave form is to approximate the saw-tooth shape sufficiently accurately for scanning purposes. Figure 8 shows the degree of approximation for a saw-tooth wave having a retrace ratio of 0.15 when 5 and 15 harmonics are included. The fifteenth harmonic extends the range up to 900 cps for the vertical scanning system, and up to 250,000 cps for the horizontal scanning system. Practical scanning generators are discussed in Sec. 67.

References

Scanning and Image Analysis:

- Baldwin, M. W., Jr.: Subjective Sharpness of Simulated Television Images, *Proc. IRE*, **28**, 458, October, 1940.
- Bedford, A. V.: Figure of Merit for Television Performance, *RMA Eng.*, **2** (1), 5, November, 1937; also *RCA Rev.*, **3** (1) 36, July, 1938.
- Cawein, M.: Television Resolution as a Function of Line Structures, *Proc. IRE*, **33**, 855, December, 1945.
- Duke, V. J.: A Method and Equipment for Checking Television Scanning Linearity, *RCA Rev.*, **6**, 190, October, 1941.
- Engstrom, E. W.: A Study of Television Image Characteristics, *Proc. IRE*, Part I, **21**, 1631, December, 1933; Part II, **23**, 295, April, 1935.
- Fink, D. G.: Photographic Analysis of Television Images, *Electronics*, **14**, 24, August, 1941.
- : Brightness Distortion in Television, *Proc. IRE*, **29**, 310, June, 1941.
- Graham and Reynolds: A Simple Optical Method for the Synthesis and Evaluation of Television Images, *Proc. IRE*, **34**, 18W, January, 1946.
- Jesty and Winch: Television Images—Analysis of Their Essential Qualities, *J. Telev. Soc.*, **2**, 316, December, 1937.
- Kallmann, H.: The Gradation of Television Images, *Proc. IRE*, **28**, 170, April, 1940.
- Kell, R. D., Bedford, A. V., and G. L. Fredendall: Determination of Optimum Numbers of Lines in a Television System, *RCA Rev.*, **5**, 8, July, 1940.
- Kell, Bedford, and Trainer: Scanning Sequence and Repetition Rate of Television Images, *Proc. IRE*, **24**, 559, April, 1936.
- Mertz and Gray: Theory of Scanning and Its Relation to the Transmitted Signal in Telephotography and Television, *Bell System Tech. J.*, **13**, 464, July, 1934.
- Wheeler and Loughren: The Fine Structure of Television Images, *Proc. IRE*, **26**, 540, May, 1938.

VIDEO-SIGNAL WAVE FORM

18. Video Signal. The video signal (or "composite video signal") is the succession of electrical impulses transmitted through the television system to convey the information from the scanning agent in the camera to the scanning agent in the receiver. Three direct functions are carried out through the video signal: (1) the transmission of impulses corresponding to the brightnesses of the scanned picture elements, conveyed by the *camera signal*; (2) the blanking of the scanning agent at the receiver during the retrace motions, by the *blanking level* or *pedestal*; and (3) the synchronization of the scanning agents, by the *vertical* and *horizontal synchronization signals*. The first item of the video signal is generated in the camera, the second two in the *synchronization signal generator*. The three items are combined in the *video mixing amplifier*.

19. Envelope of the Modulated Picture-carrier Signal. When the video signal is imposed on a carrier wave, the envelope of the modulated carrier wave constitutes the video-signal wave form. Such a modulated picture carrier and the details of the envelope are shown in Fig. 9.

In the FCC standard video signal (Fig. 9) the carrier amplitude is divided by the black level (blanking level or pedestal) at 75 ± 2.5 per cent of the maximum amplitude. The amplitude region above the black level is called the *infrablack* region and is occupied by the synchronizing signals. Signal levels in this region do not produce light in the received image. The synchronizing signals (Fig. 12) are of two types: (1) horizontal signals for initiating the motion of the scanning agent along each horizontal line and (2) vertical signals for initiating the motion of the scanning agent vertically at the beginning of each field. The peak amplitude of the wave, the height of the synchronizing pulses, and the black-level amplitude are maintained constant throughout each broadcast at the values shown in Fig. 12.

20. Camera Signal. The portion of the carrier envelope extending below the black level is called the *camera signal*. The polarity of transmission in the FCC standard is negative; *i.e.*, an increase in the light on the camera plate results in a decrease in the

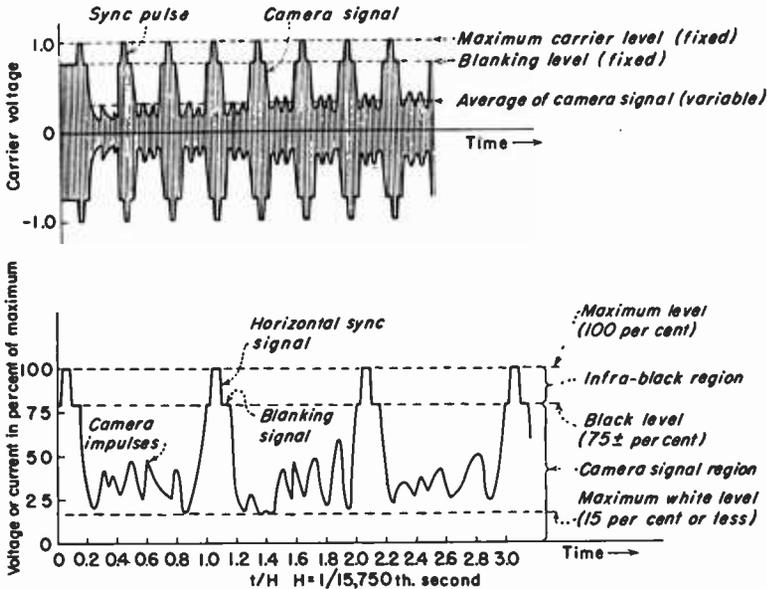


FIG. 9. (Top) Modulated television carrier signal. (Bottom) Details of modulation envelope, according to the FCC standard.

carrier amplitude, as shown in Fig. 9. The maximum white level is 15 per cent or less of the maximum carrier amplitude. Intermediate gray tones exist between the maximum white level and the black level.

The camera signal has two components (Fig. 10): an a-c component, which describes the variations in brightness from the average brightness; and the d-c component, averaged over the frame-scanning interval ($\frac{1}{30}$ sec), which represents the average or background brightness of the picture. The a-c and d-c components must be capable of being varied independently of each other, so that the same detail may be presented either on a dark background or on a bright background. Variation of the d-c component also permits the screen brightness to be "faded in" or "faded out" at the will of the studio operator. In order that the d-c component be independent of the a-c component, regardless of the changes in wave form, it is necessary that the black level be constant in the carrier envelope, and furthermore that the black level be maintained constant at the control grid of the picture tube (see Sec. 61).

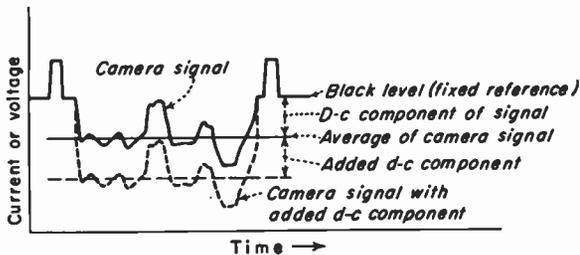


FIG. 10. Both d-c and a-c components of the modulation envelope. Dashed line represents an increase in the background brightness without change in detail.

21. Frequency Range in the Video Signal. The maximum frequency in the video range (Sec. 16) results from scanning the finest detail in the image, *i.e.*, from the scanning of adjacent picture elements. The value of f_{\max} [Eq. (13)] depends on the rate at which adjacent picture elements are scanned; values up to 4 or 5 Mc are commonly employed in present equipment.

The lowest frequency in the video range, f_{\min} , depends on the rate at which the background brightness of the scene changes. Brightness changes which take longer than

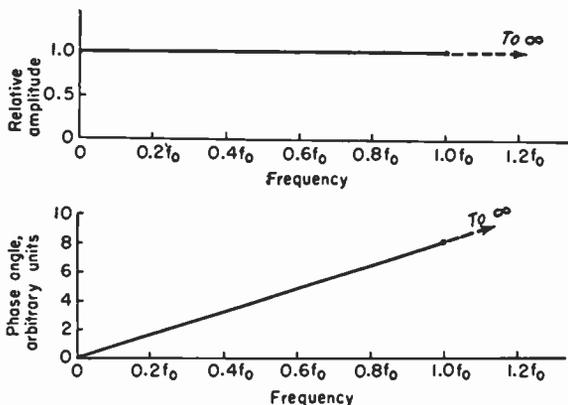


FIG. 11. Ideal phase and amplitude transmission characteristics.

the whole frame area and the area of a single picture element are reproduced by frequencies intermediate between 30 cps and 4 Mc. Since such intermediate degrees of detail may be present in any scene, the video-signal transmission system must be equally responsive to all frequencies within these limits.

22. Requirements for Transmission of the Video Wave Form. Fourier analysis of wave forms reveals that any wave form encountered in practice is composed of a number of sine-wave components having specified relative amplitudes and specified relative phases. If the wave form is to be reproduced accurately, the transmission system must be capable of transmitting all such sine-wave components, throughout the v-f range, without altering the relative amplitudes and phases of the components. This requirement is met if the amplitude-frequency response curve of the transmission system is a horizontal straight line over the v-f range, and if the phase-frequency curve of the transmission system is an oblique straight line passing through the zero-frequency point and extending over the v-f range. The ideal characteristics are shown in Fig. 11.

If the amplitude transmission characteristic is not ideal, the wave form is distorted symmetrically about a vertical axis. If the phase transmission characteristic is not ideal, the wave form is distorted asymmetrically about a vertical axis. Inadequate h-f response produces improper reproduction of steep vertical changes in the wave form. Inadequate l-f response produces improper reproduction of the flat top portions of the wave which extend over intervals comparable with the period of the low frequencies.

23. Dimensions of Synchronizing Pulses. The dimensions of the sync pulses in the FCC standard wave form are shown in Fig. 12. There are three distinct types of sync pulse. The *horizontal sync pulse* exists on the blanking pulse between the scanning of each line and occupies a duration of 8 per cent of the duration of the line-scanning interval. The *vertical sync pulse* exists on the blanking impulse between the scanning of successive fields, and extends for three times the duration of the line-scanning inter-

the duration of a single frame to complete themselves are usually introduced by changes in the d-c component of the signal. Changes that take less than the duration of a single frame are accommodated by video frequencies extending downward to 30 cps (corresponding to the frame-repetition rate of 30 per second). Consequently the significant frequency range in the video signal, based on the FCC standards, is from 30 cps to 4 Mc.

Degrees of picture detail intermediate between

val. The equalizing sync pulses exist immediately before and after the vertical sync pulse in two groups of six pulses each. The frequency of the equalizing pulses is twice that of the horizontal sync pulses.

The horizontal scanning generators at transmitter and receiver are usually synchronized by the leading edge of the horizontal pulses. Since the synchronizing action must be precise, the leading edge must be sharp. The rise of this edge must complete itself in 0.4 per cent of the line-scanning interval.

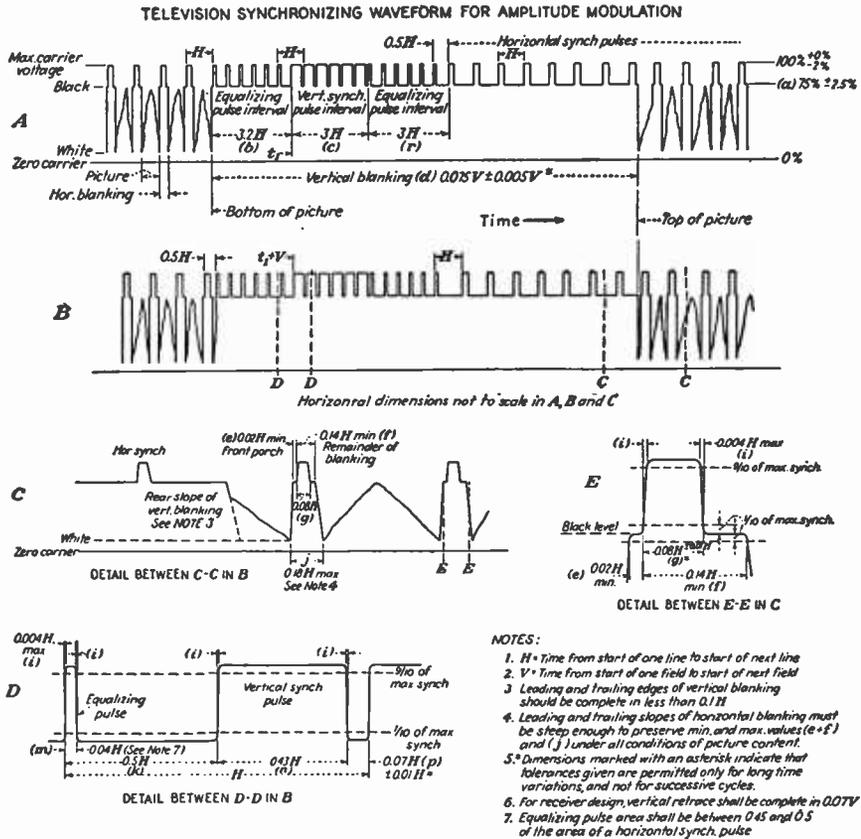


FIG. 12. The dimensions of the FCC standard video-signal waveform.

The vertical scanning generators at transmitter and receiver are usually synchronized by the integrated effect of the equalizing and vertical sync pulses which are used to charge a condenser. The leading edge of the condenser charge curve (Fig. 51) acts as the synchronizing agent. This leading edge must have precisely the same shape for each vertical pulse. The equalizing pulses are inserted to ensure that this condition is met equally for fields ending on a half line and for fields ending on a whole line.

GENERATION OF THE VIDEO-SIGNAL WAVE FORM

24. Video-signal Generator. The video-signal generator consists of three essential parts: (1) the camera and its auxiliaries, which generate the camera signal component, (2) the synchronizing signal generator, which times and shapes the vertical and hori-

zontal synchronizing signals and the blanking signals, and (3) the control amplifier which mixes the camera signals with the synchronizing signals and the blanking signals, forming the composite video signal.

25. Television Cameras. The television camera consists of a housing fitted with an adjustable camera lens which focuses the scene on the photosensitive plate of the camera tube enclosed within the housing. Also enclosed in the housing is a preamplifier which raises the level of the camera signal (usually to about 0.1 volt peak to peak)

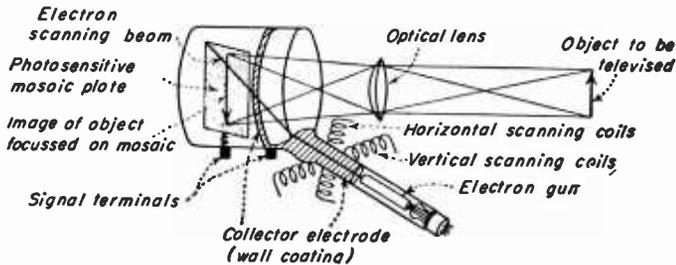


FIG. 13. Optical arrangements used with the iconoscope camera tube.

so that it can be transmitted over coaxial cable without interference. One or more scanning generators or scanning amplifiers may also be included in the camera housing. The camera is ordinarily mounted on a flexible standard so that it may be moved readily, and a universal mounting is provided so that the camera may be directed at any angle. Some form of auxiliary monitor system is also provided to enable the operator to keep the image in focus.

At present, there are four important types of electronic camera tubes: (1) the iconoscope, (2) the orthiconoscope (orthicon), (3) the image orthicon, and (4) the image dissector. The first three employ the storage principle whereby the effect of the incident light is stored as charge across a capacitance element. The image dissector is an instantaneous device, using only that light present on each picture element at the instant it is scanned. The storage devices display a luminous sensitivity from 10,000 to 100,000 times that of the instantaneous devices, depending on the storage and photoelectrical efficiencies and the number of reproducible picture elements (Sec. 12).

26. Iconoscope. A typical iconoscope is shown in Figs. 13 and 14 with its optical and electrical auxiliaries. The image is focused on the mosaic plate, which is a mica sheet coated with several million globules of photosensitized silver, insulated from each other and from a metallic coating on the reverse side of the plate. The optical image releases electrons from the mosaic, thereby charging the plate positively with a charge distribution corresponding point for point with the distribution of light in the image. The insulation prevents redistribution of this charge and permits the charge image to increase in magnitude as the light continues to fall on the mosaic.

The mosaic is scanned by a beam of electrons generated in the electron gun in the side arm of the tube. The beam, impinging on the mosaic, releases secondary electrons. The number of secondary electrons released from a given point of the mosaic depends on the potential of that point, which in turn depends on the previous photo-

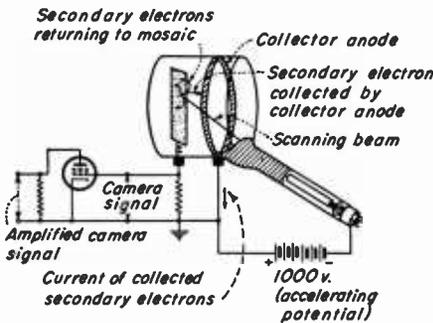


FIG. 14. Electrical arrangements used with the iconoscope camera tube.

electric emission from that point. Consequently, as the scanning agent passes over the mosaic, it generates a secondary emission current which corresponds to the successive values of brightness in the picture elements. The secondary emission is small for brightly illuminated portions of the mosaic; consequently the output current is "negative" with respect to the illumination responsible for it.

The secondary emission is collected by a collector anode and conducted through an external coupling resistor back to the metallic signal plate on the back of the mosaic support. The series circuit through which the electron current passes is accordingly composed of the ohmic resistance of the secondary emission path, the coupling resistance, and the capacitance between the signal plate and the group of globules under the scanning agent. No d.c. can flow through the capacitance; hence the output consists simply of the a-c component of the camera signal. The d-c component must be evaluated either by visual observation or by a phototube which integrates the light on the scene. A direct voltage derived from a manual control (or from the phototube in the second case) is inserted in series with the output of the iconoscope.

Since the mosaic is insulated, the current flowing toward or away from it must be zero, when averaged over any extended period of time. The average d-c value of the collected secondary emission must accordingly be replaced by electrons from the scanning beam.

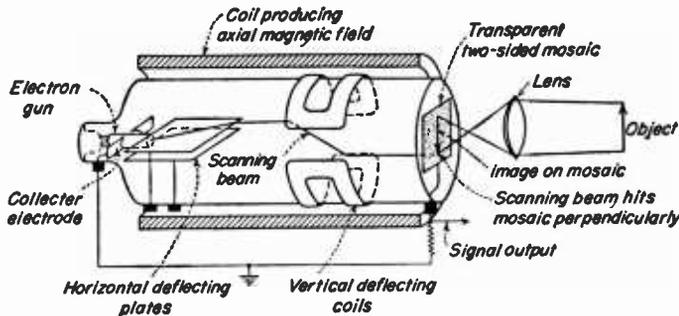


FIG. 15. Structure of the orthiconoscope ("orthicon").

Only a part of the secondary emission is collected from the mosaic. The remainder, falling back on the mosaic, sets up a distribution of charge which, when scanned, produces a spurious signal whose effect is to produce an unevenness in the background shading of the reproduced picture. This spurious signal ("dark-spot signal") must be compensated by a *shading-correction* signal generator (Sec. 35).

The color response of the typical iconoscope mosaic (when the mosaic has been silver sensitized) is very similar to that of the usual panchromatic negative-film emulsion used in motion pictures.

The sensitivity of modern iconoscopes under optimum conditions varies from about 1 mv per millilumen per sq cm illumination on the mosaic surface (low values of illumination) to about 0.25 mv (at higher illumination). The log-log curve between input illumination and output voltage of the typical iconoscope is characterized by an average slope (Sec. 72) of about 0.7. The output voltages may be increased by increasing the current used in the scanning beam, as well as by illuminating the interior of the tube envelope by a bias light.

27. Orthiconoscope. The orthiconoscope (Fig. 15) operates similarly to the iconoscope except that low-velocity electrons are used for scanning. Consequently no observable secondary emission effects arise, and no spurious "dark-spot" signal is generated. The scanning electrons themselves are collected and passed through the coupling resistor back to the mosaic. A two-sided mosaic is used.

The photoelectric emission from the mosaic is saturated in the orthiconoscope; consequently the log-log relationship between input illumination and output voltage is

linear. The sensitivity of current models is about 2 mv per millilumen per sq cm on the mosaic, although theoretical sensitivities as high as 10 mv are possible.

To use low-velocity electrons for scanning without incurring defocusing of the beam, it is necessary that the scanning beam impinge perpendicularly on the mosaic at all points in the scanning pattern. This requirement is met by a rather unorthodox deflection technique which employs a combination of axial magnetic field and transverse electric field for horizontal scanning and a transverse magnetic field for vertical scanning.

28. Image Orthicon. The most sensitive television camera tube is the image orthicon, which combines the principles of the iconoscope and the orthicon with the image dissector, including the latter's use of electron multiplication. A diagram of it is shown in Fig. 16. The optical image is focused on the *photocathode*, just inside the glass envelope of the tube. This is a continuous surface of photosensitive material which emits electrons from its rear face in a distribution proportional to the lights and shadows of the image.

The photocathode is several hundred volts negative with respect to the *target electrode* (mosaic), which is located parallel to the photocathode and about $1\frac{1}{2}$ in. distant. The electron image formed at the photocathode is accordingly drawn to the target. There the impinging electrons give rise to secondary emission, the secondaries being collected by a fine-mesh screen located adjacent to the surface of the target. The

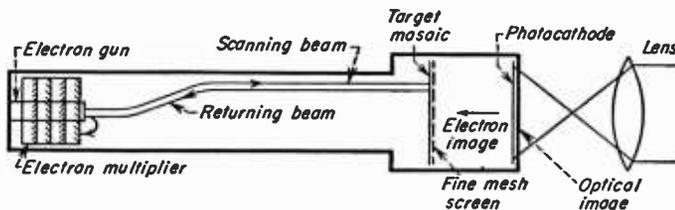


FIG. 16. Structure of image orthicon.

target thereby becomes charged with a distribution proportional to the brightnesses of the optical image, and the value of this charge is several times greater than the impinging electron image, by virtue of the electron multiplication which occurs in the act of secondary emission at the target surface.

The target is made of very thin glass having low resistivity. The lateral resistance is sufficient to preserve the charge configuration on the target for the duration of the frame interval. The target is scanned on its rear surface by a low-velocity beam of electrons in a manner identical to the scanning in the standard orthicon. Electrons are collected from the scanning beam by virtue of the positive-charge image on the reverse side of the target, and the variations thus imposed on the returning beam of scanning electrons constitute the video signal.

The electrons absorbed on the rear face of the target move through the target, by virtue of the low-resistance path offered by the glass. The charge image trapped on the forward face is thus neutralized and leaves the target prepared for the next scanning cycle.

The returning beam impinges on a target structure surrounding the electron gun, which leads to an electron multiplier. The amplified video current is collected at the output electrode of this structure and passed through the coupling resistor.

The image orthicon has a sensitivity several hundred times greater than that of the standard orthicon, the limit of sensitivity being shot-noise in the electron emission at the photocathode and the target electrode. At threshold values of light, the image orthicon exposure index exceeds that of the fastest photographic film by a factor of 10; a recognizable image may be televised with it in the light of a single candle 6 ft from the subject. The video signal is linearly related to the incident light up to a limiting value, at which a sharp saturation occurs. This permits the camera to be exposed to

widely varying levels of light without blocking or desensitizing the mosaic. The camera is widely used for televising outside events and is particularly successful under marginal conditions of light.

29. Image Dissector. The image dissector (Fig. 17) is used principally for the televising of motion-picture film, where the light source may be brilliant and highly concentrated.

The image dissector consists of a cylindrical envelope with an optical window at one end through which the image is admitted to the photocathode at the opposite end. Here an electron image is generated and drawn to the opposite end of the tube where it is focused in the plane of the scanning aperture. The aperture is fixed on the end of a finger support. The image is moved past the scanning aperture by transverse magnetic fields applied from coils external to the tube.

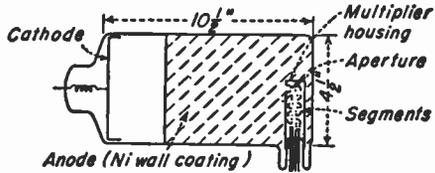


FIG. 17. Structure of the image dissector.

Inside the finger an electron multiplier structure is employed to increase the sensitivity of the device before the signal current (composed of the electrons entering the aperture) is applied to the coupling resistor. With this amplification, the signal/noise ratio of the output current is 10:1 when the mosaic illumination is 200 foot-candles. The sensitivity, when used with an 11-stage multiplier is about 50 μ v

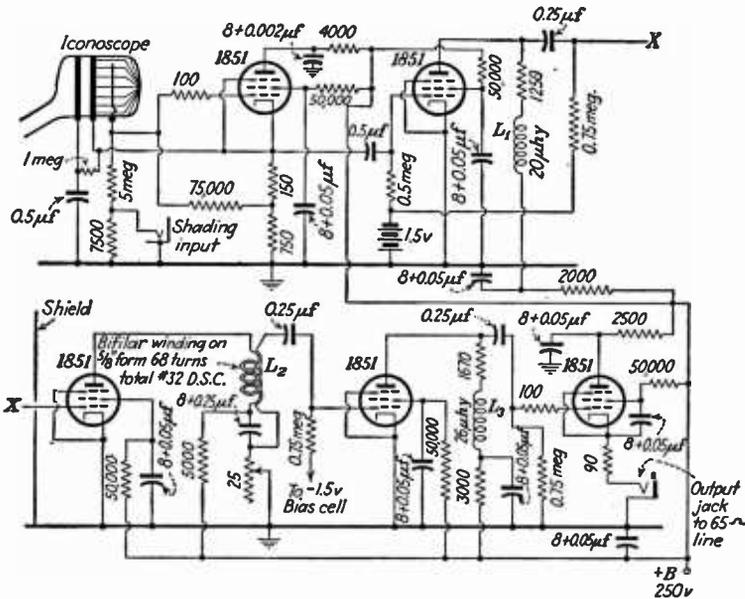


FIG. 18. Circuit diagram of an iconoscope preamplifier. (After Barco.)

per millilumen per sq cm on the photocathode, at a signal/noise ratio of 5:1. No secondary emission effects are observed. The log-log output-input curve is linear. Also the output current contains a d-c component which is directly proportional to the average brightness of the scene. Hence no auxiliary evaluation of the d-c component is necessary.

30. Preamplifier. A typical preamplifier for use with an iconoscope is shown in Fig. 18. To preserve a high signal/noise ratio in the first stage, an effective value of

successive tubes being employed to obtain the required duration, shape, and steepness of front required for these pulses. The middle chain of tubes produces the serrated vertical sync pulses continuously at 31,500 cps. The bottom chain produces equalizing pulses at 31,500 cps. All three types of pulse are produced continuously in this portion of the generator. They are interspersed in the proper order (Fig. 12) by the action of keying signals in the 6F7 tubes at the right of each chain. The interspersed signal (composite synchronizing signal) is then amplified by the stages at the extreme right and is applied to the control amplifier for mixing with the camera signal.

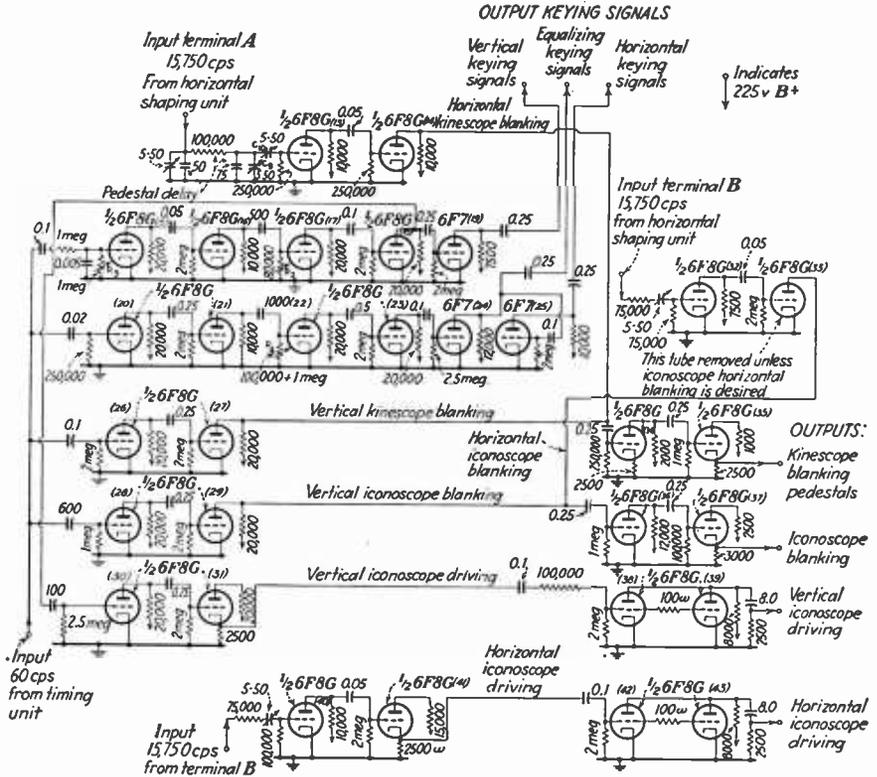


FIG. 21. Typical vertical shaping unit of a synchronization signal generator (525-line image). (After Deal.)

33. Synchronization Signal Generator. Vertical Shaping Unit. The vertical shaping unit (typical example shown in Fig. 21) has the function of producing so-called *keying signals* at a frequency of 60 cps. These keying signals are applied to the screen grids of the keying tubes in the horizontal shaping unit. The action of the keying signals is to allow to pass, or to block, the synchronizing signals passing through the keying tubes. Thus a negative keying signal is required to block the passage of the horizontal synchronizing signals during the vertical blanking period (Fig. 12); a positive keying signal is required to allow the serrated vertical sync pulse to pass at the proper time during the field blanking interval; and a two-part positive keying signal is needed to allow the equalizing pulses to pass immediately before and after the vertical sync pulses. The shape of the keying signals and the synthesis of the composite synchronizing signal are shown in Fig. 22. The vertical shaping unit accepts the 60

cps output of the timing unit and forms the required keying signals by several chains of shaping tubes which introduce the necessary wave shaping and delaying actions.

The vertical shaping unit also provides blanking signals which are applied to the control amplifier to introduce the black level during the retrace periods. Two sets of blanking signals are generated: one for the composite video signal and another, of somewhat shorter duration, for the control of the scanning beam in the camera tube. Each group of blanking signals consists of horizontal and vertical square waves recurring at 15,750 and 60 cps, respectively.

The camera-tube scanning generators are controlled by vertical and horizontal driving impulses, which are somewhat narrower and sharper than the corresponding sync pulses in the composite video signal. These driving pulses are formed from 15,750 and 60 cps signals derived from the horizontal and vertical shaping units, respectively.

34. Mixing Amplifier. The mixing amplifier (Fig. 23) has three input terminals which accept the camera signal from the camera preamplifier, the blanking signals from the synchronizing generator, and the composite synchronizing signals from the synchronizing generator. The camera signal and blanking signals are first combined by

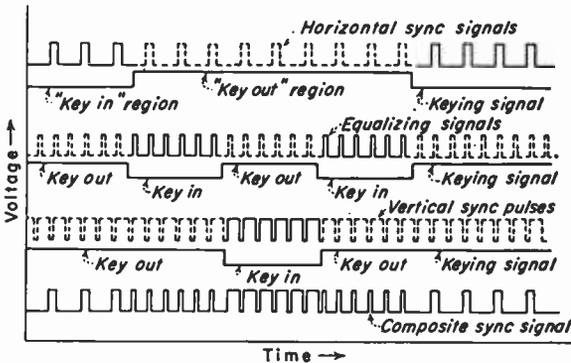


FIG. 22. Function of the keying signals in interspersing the components of the composite synchronisation signal.

means of two amplifier tubes feeding a common load resistor, across which the "semi-composite" signal appears. The d-c component of the video signal is controlled by varying the bias on the blanking signal amplifier tube, thus controlling the amplitude relationship between the average of the camera-signal component and the blanking level.

The composite synchronizing signal is similarly added to the camera and blanking components in two amplifier stages feeding a common load resistor, across which the composite video signal appears. Bias controls across these tubes control the relative amplitude of the camera and synchronizing signal amplitudes, thus allow the establishment of the 75-25 per cent relationship demanded by the standard signal.

35. Shading-correction Generator. The shading-correction generator is a device for producing wave shapes of saw-tooth, sine, and parabolic shape at vertical scanning and horizontal scanning rates (60 and 15,750 cps, respectively) in synchronism with the scanning motion. These wave shapes, controlled as to amplitude, phase, and polarity, are introduced in the preamplifier (Fig. 18) to compensate the spurious shading signal generated in the iconoscope. A form of shading-correction generator is shown in Fig. 24. The horizontal saw-tooth generator used to deflect the beam in the iconoscope tube is used directly to produce saw teeth of controlled amplitude and polarity, as well as 15,750 and 31,500 cps sine waves of controllable amplitude, phase, and polarity. Similarly the output of the vertical saw-tooth generators is used to produce similar signals. Reversal of polarity is provided by an amplifier stage. The

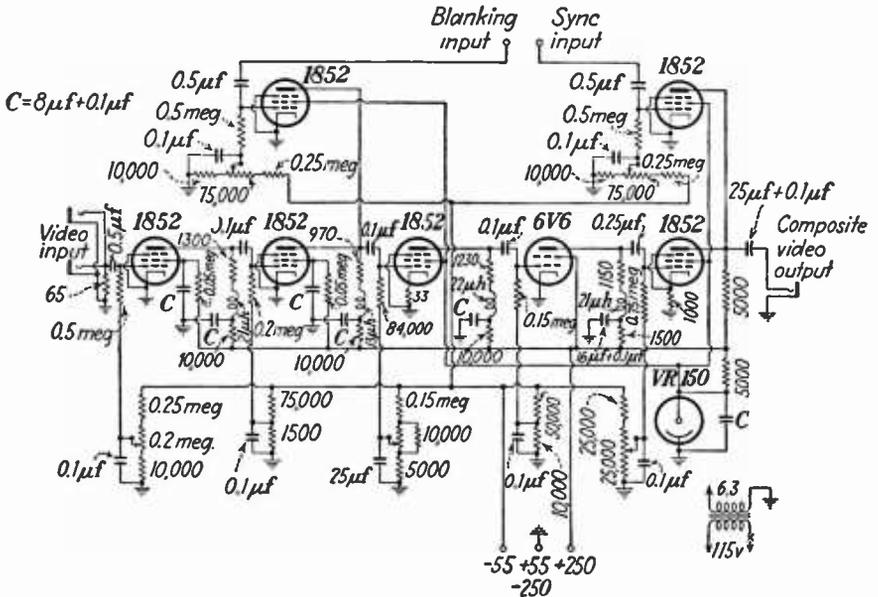


FIG. 23. Typical mixing video amplifier for combining camera signal, blanking signals, and composite synchronization signals. (After Barco.)

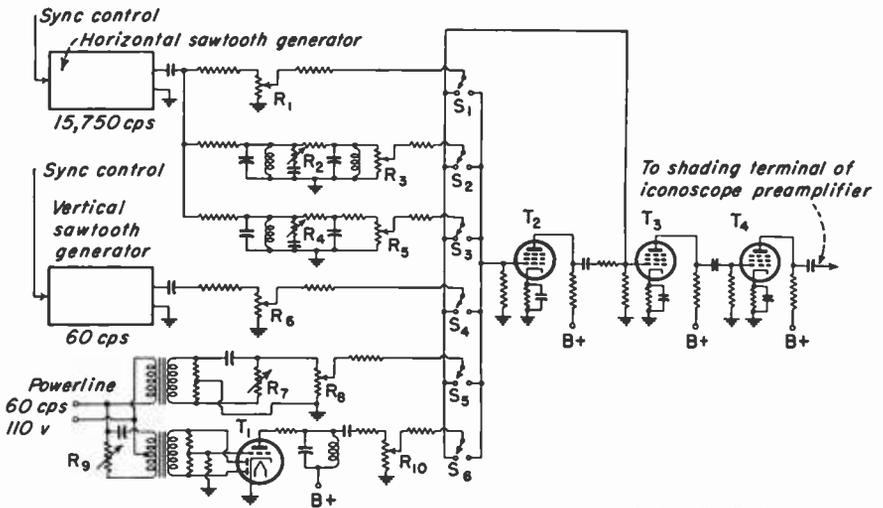


FIG. 24. Typical shading-correction signal generator. (After Bedford.)

switches control the polarity, resistors $R_1, R_2, R_5, R_6, R_8,$ and R_{10} control amplitude, and resistors $R_3, R_4, R_7,$ and R_9 control the phase. Methods of producing saw-tooth waves of controllable phase are also available, using "clipped-off" portions of the basic saw-tooth waves. The shading-correction generator controls are manipulated manually to correct for the observed defects of shading in the image as viewed on the monitor picture tube.

References

Video Signal and Its Generation, Including Camera Tubes:

- Applegarth, A. R.: Synchronizing Generators for Electronic Television, *Proc. IRE*, **34**, 128W, March, 1946.
- Barco, A. A.: Iconoscope Preamplifier Report LB-448 of the RCA License Laboratory. Information made available by special permission. See also *RCA Rev.*, **4** (1), 89, July, 1939.
- Farnsworth, P. T.: Television by Electron Image Scanning, *J. Franklin Inst.*, **218**, 411, October, 1934.
- Iams, Janes, and Hickok: The Brightness of Outdoor Scenes and Its Relation to Television Transmission, *Proc. IRE*, **25**, 1034, August, 1937.
- and Rose: Television Pick-up Tubes Employing Cathode-ray Beam Scanning, *Proc. IRE*, **25**, 1048, August, 1937.
- and ———: Television Pickup Tubes Using Low-velocity Electron Beam Scanning, *Proc. IRE*, **27**, 547, September, 1939.
- Janes and Hickok: Recent Improvements in the Design and Characteristics of the Iconoscope, *Proc. IRE*, **27**, 535, September, 1939.
- Kell, R. D., and G. C. Saiklai: Image Orthicon Camera, *RCA Rev.*, **7**, 67, March, 1946.
- Larson and Gardner: The Image-dissector, *Electronics*, **12** (10), 24, October, 1939.
- McIlwain, Knox: Survey of Television Pickup Devices, *J. Applied Phys.*, **10**, 432, July, 1939.
- Maloff, I. G.: Gamma and Range in Television, *RCA Rev.*, **3** (4), 409, April, 1939.
- Montfort and Somers: Measurements of the Slope and Duration of Television Synchronizing Pulses, *RCA Rev.*, **6**, 370, January, 1942.
- Rose, A.: The Relative Sensitivities of Television Pickup Tubes, Photographic Film, and the Human Eye, *Proc. IRE*, **30**, 293, June, 1942.
- and Iams: The Orthicon, Television Pickup Tube, *RCA Rev.*, **4** (2), 186, October, 1939.
- , Weimer, and Law: The Image Orthicon—A Sensitive Television Pickup Tube, *Proc. IRE*, **34**, 424, July, 1946.
- Zworykin, V. K.: The Iconoscope, a New Version of the Electric Eye, *Proc. IRE*, **22**, 16, January, 1934.
- , Morton, and Flory: Theory and Performance of the Iconoscope, *Proc. IRE*, **25**, 1071, August, 1937.

VIDEO AMPLIFICATION

36. Requirements for Video Amplification. The transmission system must transmit all sine-wave components within the video range, *e.g.*, 30 cps to 4 Mc, without amplitude discrimination and without phase discrimination. The gain G of a pentode amplifier stage (plate resistance large compared with the load resistance) is

$$G = g_m Z_o \quad (14)$$

where g_m is the grid-plate transconductance of the tube, and Z_o is the output impedance of the coupling connection between the stage and the following transducer. Over the video range g_m is independent of frequency; hence the amplitude and phase responses of the amplifier are determined solely by Z_o .

In video amplifiers, Z_o consists of $R, L,$ and C components so proportioned as to display a constant magnitude of impedance and a phase angle proportional to frequency over the video range. The lower frequency limit over which these conditions may be met is determined by the series impedance of the coupling capacitor, whereas the h-f limit is determined by the shunt impedance of the capacitance existing in shunt across the coupling connection. The value of gain within these limits depends on g_m and on the value of the load resistor, since this is the principal component of Z_o within the video range.

37. H-f Compensation. To minimize the effect of the shunt capacitance, it is usual in video amplifiers to insert a small inductance either in series with the load resistor (shunt peaking), in series with the coupling connection (series peaking), or a combination of the two (shunt-series peaking). The inductance is used to form a resonant circuit with the shunt capacitance at a frequency above the upper limit of the required v-f range, and the rising resonance characteristic is used to counteract the falling off

of the Z_0 value at the upper frequency limit. The load resistor must similarly be chosen in terms of the total shunt capacitance, so that the gain in the mid-frequency range (where reactive effects are not prominent) will be the same as at the upper limit (where reactive effects are predominant).

In all cases of h-f compensation the basic factor is the total shunt capacitance C_i associated with the coupling connection

$$C_i = C_{pk} + C_{pk} + C_{op}(G + 1) + C_{stray} \tag{15}$$

where C_{pk} = output tube capacitance

C_{pk} = input capacitance of following tube

C_{op} = grid-plate capacitance of following tube

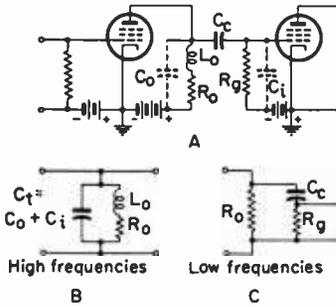
G = stage gain of following stage

C_{stray} = total shunt capacitance due to wiring, tube sockets, terminals, etc.

In pentode amplifiers, C_{op} may ordinarily be neglected.

38. Shunt-peaking Compensation.

The most widely used h-f compensation scheme (Fig. 25) is known as *shunt peaking*, because the resonating (peaking) inductance L_0 is in shunt across the shunt capacitance C_i . The design values of L_0 and R_0 (the load resistor) are based on the shunt capacitance C_i , on the maximum required frequency in the video range f_{max} and on two design constants k_L and k_R which relate the impedance of L_0 and R_0 , respectively, to the impedance of C_i at the maximum frequency f_{max} .



$$k_R = \frac{R_0}{1/(2\pi f_{max} C_i)} \tag{16}$$

$$k_L = \frac{2\pi f_{max} L_0}{1/(2\pi f_{max} C_i)} \tag{17}$$

FIG. 25. High-frequency compensation by the shunt-peaking method, with equivalent circuits for high and low frequencies.

The values of k_R range from 0.8 to 1.0; most designs are based on $k_R = 1$; i.e., the load resistance is made equal to the impedance of C_i at the maximum v.f. The values of k_L range from 0.3 to 0.7, but most designs are based on $k_L = 0.5$; impedance of the inductance L_0 is made one-half as great as the impedance of C_i , at f_{max} . This is equivalent to making the resonant frequency between L_0 and C_i equal to 1.41 times f_{max} .

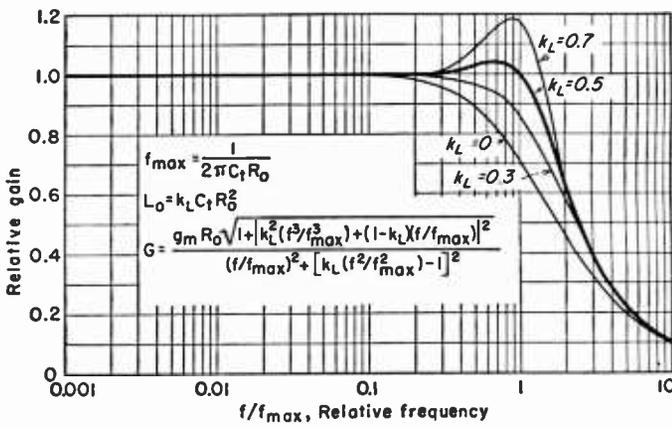


FIG. 26. Gain of a shunt-compensated video amplifier.

On the assumption that $k_R = 1.0$, the expression for the gain of the shunt-compensated video amplifier is

$$G = \frac{g_m R_0 \{1 - j[k_L^2 (f/f_{max})^3 + (1 - k_L) (f/f_{max})]\}}{(f/f_{max})^2 + [k_L (f/f_{max})^2 - 1]^2} \tag{18}$$

where G is the gain at frequency f , and the other quantities have been defined. The absolute magnitude of this equation is plotted in Fig. 26, and its phase angle in Fig. 27, for several values of k_L .

The simplified design equations for shunt peaking ($k_R = 1$ and $k_L = 0.5$) are as follows:

$$R_0 = \frac{1}{2\pi f_{max} C_i} \tag{19}$$

$$L_0 = 0.5 C_i R_0^2 \tag{20}$$

Typical values of R_0 are 2,000 to 4,000 ohms and of L_0 are 50 to 100 μ h.

39. Series-peaking Compensation. The compensation in Fig. 28 has an advantage over the shunt-peaking system in that the inductance L_c isolates the effects of the output and input capacitances C_0 and C_i ; whereas in the shunt-peaking systems, C_0 and C_i are directly additive. Since C_0 is less than C_i for a given h-f limit R_0 may be made correspondingly larger; hence the gain of the stage is increased. On the assumption

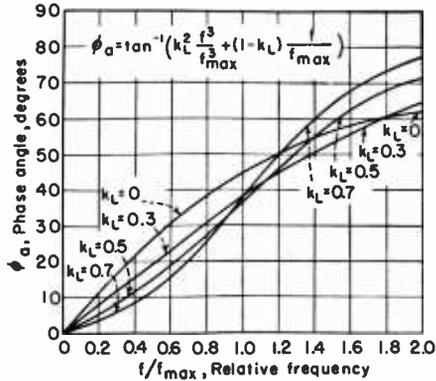


FIG. 27. Phase angle introduced by a shunt-compensated video amplifier.

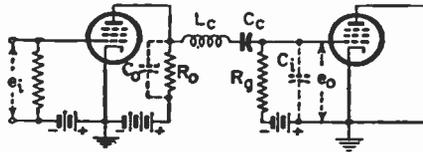


FIG. 28. The series-peaking system of high-frequency compensation.

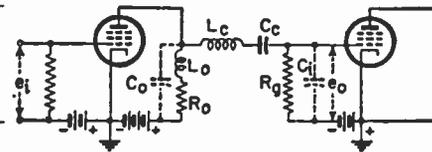


FIG. 29. Combined shunt- and series-peaking system of high-frequency compensation.

that $C_i/C_0 = 2$ (usually assumed condition), the design equations for R_0 and L_c are as follows:

$$R_0 = \frac{1.5}{2\pi f_{max} (C_0 + C_i)} \tag{21}$$

$$L_c = 0.67 C_i R_0^2 \tag{22}$$

With these values the gain is uniform up to f_{max} , and its value is 50 per cent greater than the gain of the shunt-compensated stage with the same values of C_0 , C_i , and C_i , provided $C_i/C_0 = 2$.

40. Shunt-series-peaking Compensation. The combination of shunt and series peaking (shown in Fig. 29) allows still higher gain by combining the virtues of both connections. Assuming $C_i/C_0 = 2$, the design equations are

$$R_0 = \frac{1.8}{2\pi f_{max} C_i} \tag{23}$$

$$L_0 = 0.12 C_i R_0^2 \tag{24}$$

$$L_c = 0.52 C_i R_0^2 \tag{25}$$

The stage displays up to f_{max} uniform gain, which is 80 per cent greater than that of

the simple shunt-peaking stage. The relative merits and design factors of the three methods of h-f compensation are shown in Table 2.

Table 2. High-frequency Compensation Systems

Type	R_0	L_0	L_c	Relative gain at f_{max}	Variation in time delay, sec up to f_{max} cps
Uncompensated.....	$1/(2\pi f_{max} C_i)$	0.707	$0.035/f_{max}$
Shunt.....	$1/(2\pi f_{max} C_i)$	$0.5C_i R_0^2$	1.0	$0.023/f_{max}$
Series ($C_i/C_0 = 2$).....	$1.5/(2\pi f_{max} C_i)$	$0.67C_i R_0^2$	1.5	$0.0113/f_{max}$
Shunt-series ($C_i/C_0 = 2$).....	$1.8/(2\pi f_{max} C_i)$	$0.12C_i R_0^2$	$0.52C_i R_0^2$	1.8	$0.015/f_{max}$

41. L-f Compensation. The amplitude response of conventional resistance-capacitance-coupled amplifier stages at low frequencies is usually satisfactory, but the phase response at the low frequencies is troublesome.

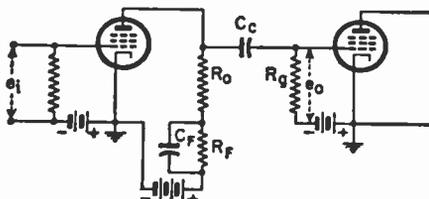


FIG. 30. RC method of low-frequency compensation.

The phase angle introduced by the coupling connection C_c and the grid resistor R_g of the following stage is sufficient to prevent proper reproduction of square waves of 30 or 60 cps fundamental frequency, unless very large values of C_c and R_c are employed. Large values of C_c introduce shunt capacitance to ground, and large values of R_g introduce grid-current difficulties in the following stage. Large values of $C_c R_g$ may induce relaxation oscillations.

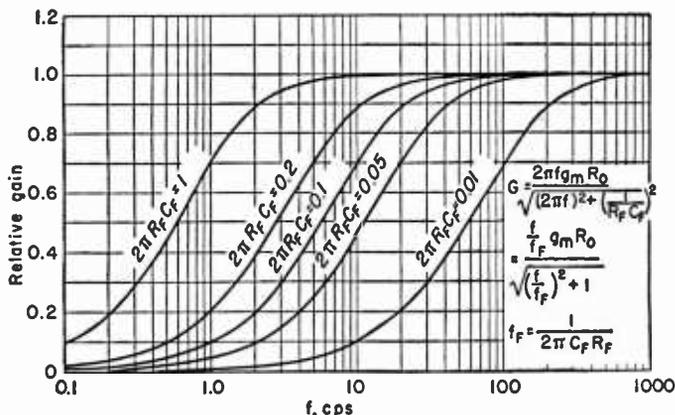


FIG. 31. Amplitude response of low-frequency compensation system.

Accordingly it is usual to compensate the effect of the time constant $C_c R_g$ by the introduction of a filter $R_F C_F$ shown in Fig. 30.

$$\frac{C_F R_0 R_F}{R_0 + R_F} = C_c R_g \tag{26}$$

When this condition is met, the gain at low frequencies is

$$G = \frac{(f/f_F)g_m R_0}{f/f_F - j} \tag{27}$$

where G = gain at frequency f
 $f_F = 1/(2\pi C_F R_F)$
 $j = \sqrt{-1}$

The amplitude and phase of Eq. (27) are shown in Figs. 31 and 32. Values of $R_F C_F$ from 0.15 to 0.5 should be used to keep the point of zero-phase shift below 30 cps, as indicated in Fig. 32.

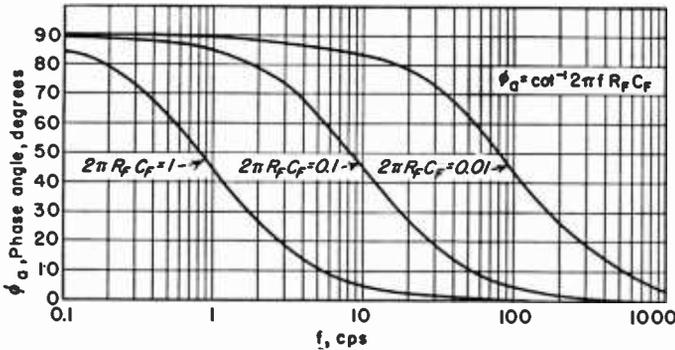


FIG. 32. Phase response of low-frequency compensation system.

42. Cathode-coupled Stage. For many purposes a video-amplifier stage displaying low output impedance is necessary (to match the impedance of coaxial cables and to permit the stage to feed many high impedance sources at once). The cathode-coupled stage (Fig. 33) is commonly used for this purpose. The gain of this stage is less than unity, and its output impedance can be designed readily for values as low as 50 ohms. The amplifier, being degenerative, has lower values of input capacitance, is freer from amplitude distortion, and is less affected by changes in supply voltages than is the conventional amplifier stage.

The gain of the cathode-coupled stage is

$$G = \frac{\mu R_k}{r_p + R_k(\mu + 1)} \tag{28}$$

where μ = amplification factor of tube
 r_p = its internal plate resistance
 R_k = value of cathode resistor

The effective output impedance Z_o is

$$Z_o = \frac{R_k r_p / (\mu + 1)}{R_k + r_p / (\mu + 1)} \tag{29}$$

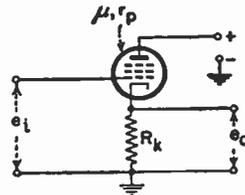


FIG. 33. Fundamental cathode-coupled stage.

An important practical advantage of the cathode-coupled stage is that it may be coupled to the following transducer without the intervention of a coupling capacitor, so that the d-c as well as a-c components of the video signal are transmitted. No pains need be taken to preserve the h-f response, since the low value of impedance makes the shunting effect of the output capacitance negligibly small.

43. Transient Response of Video Amplifiers. The response of a video amplifier to the Heaviside unit pulse of voltage is a general criterion of video-amplifier response. The response to a single unit pulse is difficult to measure experimentally, but a square wave may be used as the exciting voltage, provided that the period of the wave is long compared with the duration of the transient response. Responses calculated on this assumption are shown in Fig. 34, for a single stage and for several multistage amplifiers. Simple shunt peaking is assumed, for various values of the

parameter $K = RC\sqrt{LC}$, i.e., the ratio of the load resistor R_0 to the impedance of the shunt capacitance C_i at the frequency at which L_0 and C_i are resonant. The case for $K = 1.41$ is equivalent to the cases of $k_R = 1$ and $k_L = 0.5$ (Sec. 38).

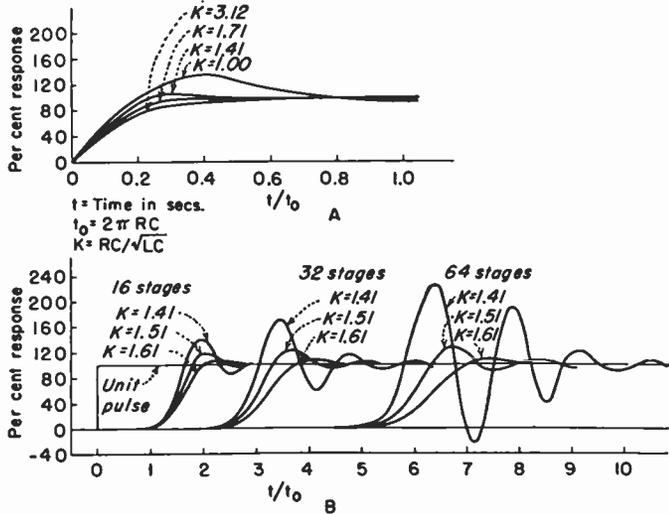


FIG. 34. Transient response of single and multistage compensated video amplifiers. (After Bedford and Fredenhall.)

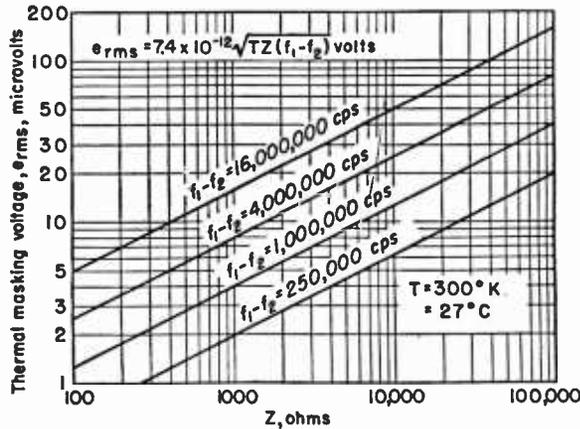


FIG. 35. Thermal agitation voltage generated in wide-band circuits.

44. Noise Limitations to Video Amplification. One of the principal limitations to proper video amplification is inadequate signal/noise ratio. The two sources of circuit noise, thermal agitation and shot effect, are evaluated in Figs. 35 and 36 for a transmission system responsive to the video range. Values of 50 to 100 μv are common. For a signal/noise ratio of 10:1, commonly assumed as the minimum acceptable for entertainment purposes, the desired signal must accordingly have an rms amplitude of from 0.5 to 1.0 mv.

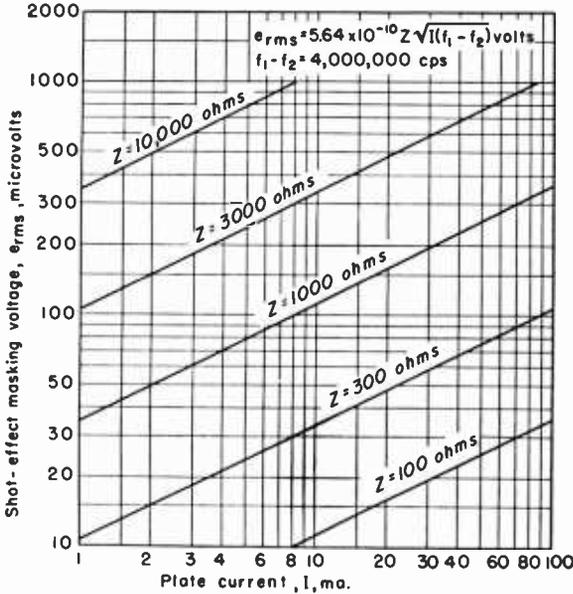


FIG. 36. Shot-effect noise generated in wide-band circuits.

References

Video Amplification:

Barco, A. A.: Measurement of Phase Shift in Television Amplifiers. *RCA Rev.*, 3 (4), 441, April, 1939.

Bedford and Fredenhall: Transient Response of Multistage Video Amplifiers, *Proc. IRE*, 27, 277, April, 1939.

Bereskin, A. B.: Improve High-frequency Compensation for Wideband Amplifiers, *Proc. IRE*, 32, 608, October, 1944.

Builder, G.: The Amplification of Transients, *Wireless Engr., Exptl. Wireless*, 246, May, 1935.

Carnahan, C. W.: The Steady-state Response of a Network to a Periodic Driving Force of Arbitrary Shape, and Its Applications to Television Circuits, *Proc. IRE*, 23, 1393, November, 1935.

Everest, E. A.: Wideband Television Amplifiers, *Electronics*, 11 (1), 16, January, 1938; 11 (5), 24, May, 1938.

Foster and Rankin: Video Output Systems, *RCA Rev.*, 5, 409, April, 1941.

Freeman and Schantz: Video Amplifier Design, *Electronics*, 10 (8), 22, August, 1937.

Herold, E. W.: High-frequency Correction in Resistance-coupled Amplifiers, *Communications*, 18 (8), 11, August, 1938.

Kallman, Spencer, and Singer: Transient Response, *Proc. IRE*, 33, 169, March, 1945.

Larsen, M. J.: Low-frequency Compensation of Video-frequency Amplifiers, *Proc. IRE*, 33, 666, October, 1945.

Maclean, W. R.: Ultimate Bandwidths in High-gain Multistage Video Amplifiers, *Proc. IRE*, 32, 12, January, 1944.

McLachlan, H. W.: Reproduction of Transients by Television Amplifiers, *Wireless Engr.*, 13, 519, October, 1936.

Puckle, O. S.: Transient Aspect of Wideband Amplifiers, *Wireless Eng.*, 12, 251, May, 1935.

Schlesinger, K.: Low-frequency Compensation for Amplifiers, *Electronics*, 21, 103, February, 1948.

Seeley and Kimball: Analysis and Design of Video Amplifiers, *RCA Rev.*, 2 (2), 171, October, 1937; 3 (3), 290, January, 1939.

Swift, G.: Amplifier Testing by Means of Square Waves, *Communications*, 14 (2), 22, February, 1939.

Sziklai and Schroeder: Cathode-coupled Wideband Amplifiers, *Proc. IRE*, 33, 701, October, 1945.

Wheeler, H. A.: The Interpretation of Amplitude and Phase Distortion in Terms of Paired Echoes, *Proc. IRE*, 27, 359, June, 1939.

—: Wideband Amplifiers for Television, *Proc. IRE*, 27, 429, July, 1939.

MODULATION, R-F AND I-F AMPLIFICATION, DETECTION

45. Video Modulation. Video modulation is based on the same considerations as audio modulation, with certain specialized requirements. One of the limitations is the

small amount of video-signal voltage which may be generated in currently available tubes and circuits. The high capacitance to ground of large water-cooled tubes requires the use of very low values of load resistance to maintain response over the v-f range. The voltage which can be developed across the load resistance depends on the emission current. At present it is difficult to generate more than the 1,500 or 2,000

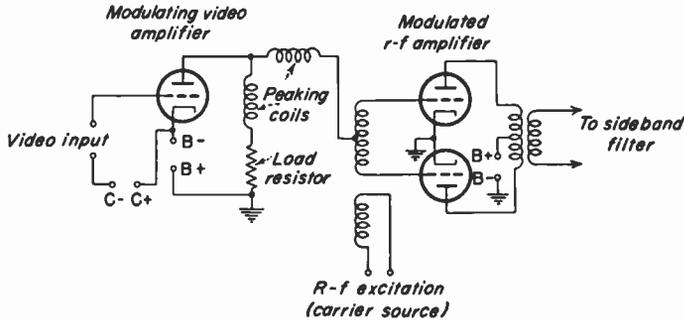


FIG. 37. Fundamental circuit of video modulator.

volts peak to peak over the video range from 30 cps to 4.5 Mc. When high-level modulation is used, therefore, it is usually considered expedient to use grid-circuit modulation, rather than plate-circuit modulation, since the voltage requirements for grid modulation are less by the amplification factor of the modulated stage. Low-level

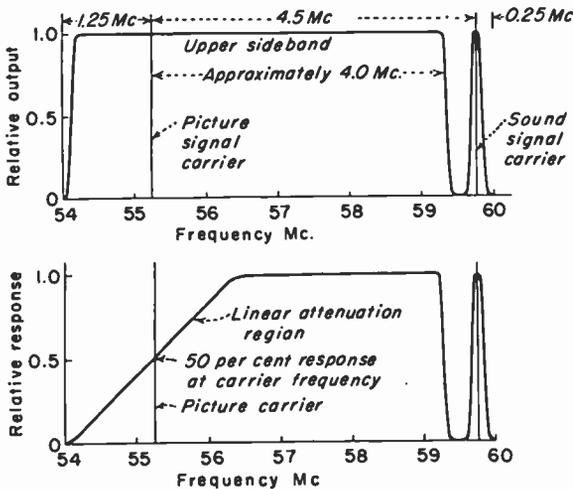


FIG. 38. (Top) Output characteristic of television transmitter. (Bottom) Corresponding response characteristic of receiver.

This makes necessary a separate power supply for each stage. A typical arrangement is shown in Fig. 37. Here the modulating video amplifier is coupled conductively to the grids of the r-f amplifier. The B supply for the modulating amplifier is in series with the cathode.

At the grid of the modulating amplifier, it is necessary that the blanking level and sync-pulse tip level be constant. The latter levels are caused to assume fixed values

modulation is not similarly restricted but requires a number of linear class B r-f amplifier stages following the modulator.

The second unusual requirement in video modulation is the necessity for maintaining two levels in the modulation envelope at constant amplitudes. These levels are (1) the tips of the sync pulses, i.e., the maximum amplitude of the envelope, and (2) the blanking level or pedestal. Since these levels must remain constant regardless of any changes in the wave form of the camera-signal component, it is necessary to couple the modulating amplifier conductively to the modulated amplifier.

by passing the video wave form through a diode rectifier whose cathode is connected to the modulating video-amplifier grid. The load circuit values are chosen so that the rectified d-c potential across the diode assumes a level at the tips of the sync pulses, or just below the tips (the difference being required to supply the diode current). The voltage across the diode forms a part of the fixed bias of the modulating amplifier. The composite wave form, extending more positively than the tips of the sync pulses, causes the modulating amplifier output voltage to extend more negatively than the sync pulses. This output voltage, applied to control the amplitude of the modulated r-f amplifier, causes the sync-pulse tips to assume the peak position in the envelope, while the blanking level and camera-signal components extend to lower levels in the envelope. The sync pulses and blanking level maintain constant amplitudes, whereas the average on the camera-signal component changes with the background illumination of the scene (see Fig. 10).

46. Vestigial-side-band Transmission. The side bands of the modulated r-f signal, assuming a maximum video modulating frequency of 4.5 Mc, extend over a total region of 9 Mc. To conserve space in the ether and at the same time to secure greater efficiency from r-f and i-f amplifiers, the FCC standards specify vestigial-side-band transmission ("single"-side-band transmission). In this system a part of the lower frequency side band is completely attenuated. By this means the upper side band can be transmitted completely with 4.0 to 4.5 Mc width, within the 6-Mc channel assigned by the FCC. A portion of the lower side band, within 1.25 Mc of the carrier frequency, is also transmitted.

The channel composition for vestigial-side-band transmission is shown in Fig. 38, at the top. The lower figure shows the

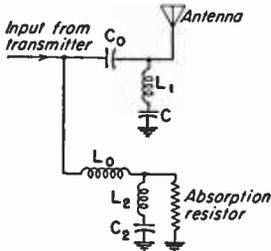


FIG. 39. Elements of a vestigial-side-band filter.

Amateur - 50-54
Channel 2-54-60
Channel 3-60-66
Channel 4-66-72
Non-Gov. 72-76
Channel 5-76-82
Channel 6-82-88
FM Broadcast 88-108
Government 108-118
Airport 118-122
Non-Gov. 122-132
Government 132-144
Amateur 144-148
Government 148-152
Non-Gov. 152-162
Government 162-174
Channel 7-174-180
Channel 8-180-186
Channel 9-186-192
Channel 10-192-198
Channel 11-198-204
Channel 12-204-210
Channel 13-210-216

FIG. 40. Allocation of frequencies from 50 to 216 Mc.

corresponding characteristic of the receiver. In the receiver characteristic the picture signal-carrier voltage is attenuated to 50 per cent of its original level, and the curve of attenuation is linear over a range of 2.5 Mc. This arrangement develops 50 per cent modulation in those portions of the carrier which receive double-side-band treatment (within 1.25 Mc of the carrier). The modulation of the components further removed from the carrier in the upper side band are inherently 50 per cent modulated, so all portions of the signal, when presented to the detector, produce an equal amplitude in the detector output.

To form a transmitted signal of the character shown at the top in Fig. 38, a filter having sharp cutoff characteristics is required. In Fig. 39 the desired upper side band is passed through a capacitor to the antenna, whereas the undesired lower side band is passed through an inductance to an absorbing resistor. Filter structures for this purpose, when employed

for high-level modulation, are customarily formed from sections of coaxial transmission lines. In addition to the filter shown, a sharply tuned "notching filter" is usually used to provide additional attenuation at the sound-carrier frequency of the adjacent channel. The characteristic shown in Fig. 38 may also be obtained with low-level modulation by proper positioning of the pass bands of the linear class B stages that follow the modulator.

47. Allocation of Television Channels. Figure 40 shows the allocation of twelve 6-Mc channels for television in the region between 54 and 216 Mc allocated by the FCC.

48. Wide-band R-f Amplification. The gain of a wide-band r-f pentode amplifier is equal to the product of the tube transconductance g_m by the load impedance Z_0 . For maximum gain per stage both g_m and the absolute value of Z_0 must be as large as possible. To maintain the proper band-pass characteristics, Z_0 must be designed to have as nearly constant amplitude and as nearly linear phase as possible over the desired operating range and to have nearly zero impedance outside these limits. The

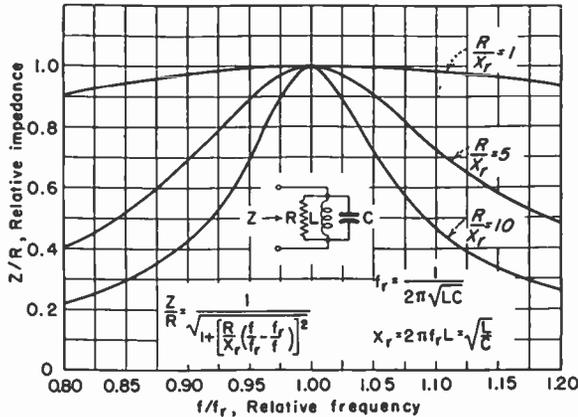


Fig. 41. Impedance characteristics (absolute magnitude) of the single loaded tuned circuit (inset).

design of the optimum Z_0 to meet these conditions is best attacked from the standpoint of band-pass filter theory.

Some general considerations are revealed simply in the analysis of the single loaded tuned circuit, shown in Fig. 41. At the resonant frequency f_r , the impedance of the circuit is equal to the value of the shunting resistor. At frequencies removed from resonance, the impedance is less by the amount shown, and the degree of attenuation depends on the ratio of the resistance to the impedance of the inductance and resistance at the resonant frequency. The general relationships between the impedance and resistance are shown in Fig. 41. The corresponding phase relationships are shown in Fig. 42.

The design of the single loaded tuned circuit is based on the necessity of (1) obtaining resonance at or near the carrier frequency and (2) loading the circuit to present nearly uniform response over the side-band regions. In particular, if it is desired that the circuit display an impedance at the edge of the side-band regions equal to 0.707 times the impedance at resonance, the value of the resistance required is

$$R = \frac{f_r}{\Delta f} \sqrt{\frac{L}{C}} \tag{30}$$

where f_r = resonant frequency

Δf = total frequency width of region within which the response is desired within unity and 0.707

R = max value of impedance which the tuned circuit impedance Z_0 may have

For maximum gain the L/C ratio should be as high as possible. It is usual to employ as C only the stray and distributed capacitance present in the circuit and to bring this capacitance to resonance by employing a variable value of L . The loading is determined from Eq. (30).

Since the single-tuned circuit cannot display a uniform impedance over an extended band width, it is usually desirable to employ coupled circuits (coupled capacitively, self-

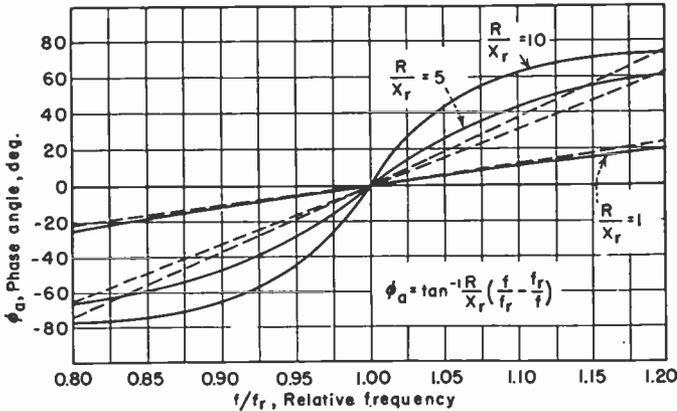


FIG. 42. Phase angle of impedance of single loaded tuned circuit.

inductively, or mutually inductively) to obtain a flattop response curve. In coupled circuits the impedance is a complicated function, but in general its value is independent of the carrier frequency, inversely proportional to the band width, and directly proportional to the L/C ratio of the tuned circuits.

49. Picture I-f Amplification. The design of i-f amplifier circuits for television i-f signals is similar to the design of wide-band r-f circuits, except that lower signal frequencies are used. Also, since most of the gain in a television receiver resides in the i-f amplifier stages, the problem of selectivity against interference from adjacent channels must be contended with in this portion of the receiver.

The band-pass characteristic of an ideal picture i-f amplifier is shown in Fig. 43. The values of carrier frequencies are those recommended by the RMA Committee on Television Receivers, i.e., 21.25 Mc for the sound carrier and 25.75 Mc for the picture. A more recent (1949) recommendation for these values is 41.25 Mc and 45.75 Mc respectively. The adjacent carrier frequencies are 27.25 Mc for the adjacent sound channel, and 19.75 Mc for the picture carrier of the oppositely adjacent channel. Ordinarily the picture i-f circuit need be designed to display selectivity against the two sound channels only, at 21.25 Mc against the associated sound channel and at 27.25 Mc against the adjacent-channel sound carrier. The usual values of attenuation are 40 db at 21.25 Mc and 60 db at 27.25 Mc.

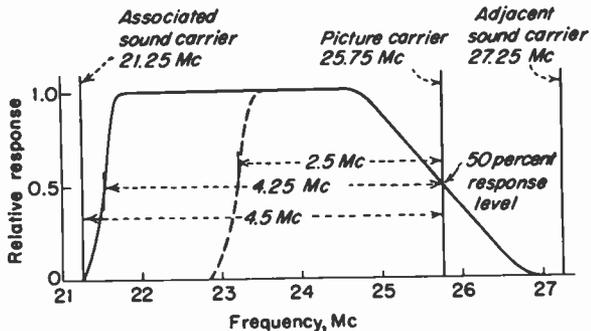


FIG. 43. Response curve of typical television i-f amplifier, according to RMA recommended practice.

To produce the recommended standard intermediate frequencies, the frequency of the local oscillator must be 21 Mc higher in frequency than the upper frequency limit of the channel under consideration. For the new i-f values the local oscillator should be 41 Mc higher.

The gain per stage in picture i-f amplifiers depends directly on the band width passed. Stage gains of 10 are possible when accepting the full band width of 4 Mc shown in Fig. 43. A total i-f gain of 10,000 is usually considered sufficient. The effective tube and circuit noise at the input to the first i-f stage is usually 100 μ v or more. With a gain of 10,000 the noise voltage applied to the detector would be 1 volt, which is sufficient to make it plainly visible in the cathode tube. Sensitivity greater than this is clearly not necessary. Total i-f gain as low as 5,000 may be used in low-priced receivers, intended for use with input r-f signals of 500 μ v or more.

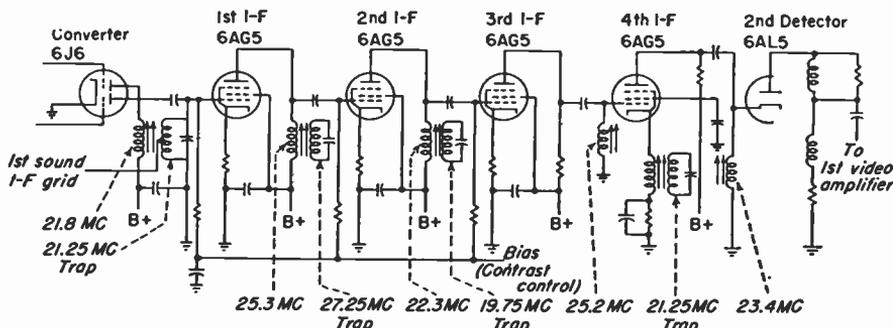


FIG. 44. Diagram of typical picture i-f amplifier.

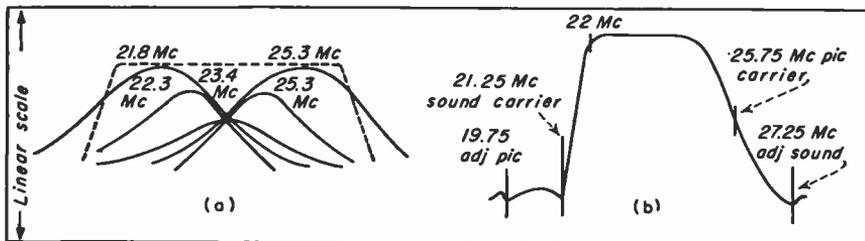


FIG. 45. Response curve of picture i-f amplifier shown in Fig. 44.

50. Video Detection. The diode detector is used almost universally for video demodulation in current receivers. The important considerations are (1) the amplitude and phase responses of the load circuit of the detector over the video range, (2) the discrimination of this circuit against components of carrier frequency, (3) the loading exerted by this circuit on the i-f coupling circuit which feeds the detector, and (4) the polarity of the detected voltage output.

In designing the detector load circuit, the important factors are the capacitance to ground of the detector output and the input capacitance of the following video amplifier. The circuit usually used is very similar to the series-peaking circuit (Sec. 39), and the expressions for R_0 and L_c [Eqs. (21) and (22)] can be used, under the assumption that $C_i/C_0 = 2$. The values of R_0 , so determined, usually range from 2,000 to 5,000 ohms.

The simple series-peaking circuit possesses sufficient discrimination against carrier-frequency components when the detection occurs at radio frequencies (above 40 Mc). But when i-f detection is considered (frequencies from 21.25 to 25.75 Mc), it is preferable to design the detector load circuit in the form of a low-pass filter having a sharp cutoff above the v-f limit (5 Mc). A typical constant- k filter section of this type is shown in Fig. 46.

The loading of the detector load circuit on the preceding i-f circuit is calculated from

$$R_{\text{eff}} = \frac{R}{2\nu} \tag{31}$$

where R_{eff} = effective load resistance on i-f amplifier

R = actual value of detector load resistor

ν = detection efficiency (very close to unity in most practical cases)

The polarity of the detected voltage output is important because it determines the number of video-amplifier stages required between the detector and the picture tube control grid to produce a picture having positive tone values. The two possible detector polarities are shown in Fig. 47. The cathode-above-ground connection produces an increased voltage output as the initial light in the studio decreases (assuming negative modulation, see Sec. 20). Consequently one phase reversal is necessary between the detector and the picture tube. Any odd number of video-amplifier stages suffices (usually one stage is used). In the anode-above-ground connection the reverse polarity exists and an even number (usually two) of stages is required between the detector and picture tube. The same polarity considerations govern the number of amplifier stages required between the detector and the synchronizing input terminals of the scanning generators. With scanning generators synchronized by positive pulses (usual type) the cathode-above-ground connection shown at A requires an even number of intervening stages, whereas the anode-above-ground connection B requires an odd number of stages.

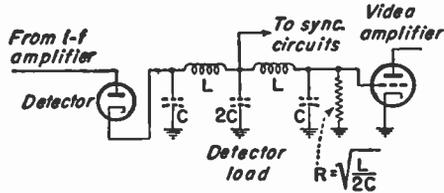


FIG. 46. Typical video detector circuit with filter load circuit.

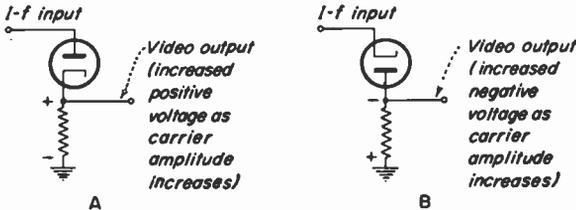


FIG. 47. Detector polarities: (A) cathode-above-ground connection; (B) anode-above-ground connection.

It is usual to operate video detectors with a maximum peak-to-peak input i-f voltage of 10 volts. Assuming full modulation, the peak-to-peak output voltage (with detector internal resistance and load resistance values equal) will be 5 volts, three-quarters of which constitutes the camera signal. A single video stage having a gain of 12 is consequently capable of delivering $5.0 \times 0.75 \times 12 = 45$ volts, peak to peak, in the picture tube grid. This value is sufficient to operate the usual picture tube over its entire control range.

References

Modulation, R-f and I-f Amplification, Detection:

Benham, W. E.: Aerial Coupling System for Television, *Wireless Engr.*, 18, 555, October, 1938.
 ———: Asymmetric Sideband Phase Distortion, *Wireless Engr.*, 18, 616, November, 1938.
 Brown, G. H.: A Vestigial Sideband Filter for Use with a Television Transmitter, *RCA Rev.*, 5, 301, January, 1941.
 Carter, P. S.: Simple Television Antennas, *RCA Rev.*, 4 (2), 168, October, 1939.
 Dome, R. B.: Carrier Difference Reception of Television Sound, *Electronics*, 20, 102, January, 1947.
 Fredendall, Schlesinger, and Schroeder: Transmission of Television Sound on the Picture Carrier, *Proc. IRE*, 34, 49P, February, 1946.
 Hollywood, J. M.: Single Side-band Filter Theory with Television Applications, *Proc. IRE*, 27, 457, July, 1939.
 Landon, V. D.: Cascade Amplifiers of Maximal Flatness, *RCA Rev.*, 5, 347, January, 1941; 5, 347, April, 1941.

Mountjoy, Garrard: Television Signal-frequency Circuit Considerations, *RCA Rev.*, 4 (2), 204, October, 1939.

Poch and Epstein: Partial Suppression of One Sideband in Television Reception, *RCA Rev.*, 1 (3), 19, July, 1937.

Seeley: Effect of Receiving Antenna on Television Reception Fidelity, *RCA Rev.*, 2 (4), 433, April, 1938.

Strutt, M. J. O.: High Frequency Mixing and Detector Stages in Television Receivers, *Wireless Engr.*, 16, 174, April, 1939.

Sziklai and Schroeder: Band-pass Bridged-T Network for Television Intermediate Frequency Amplifiers, *Proc. IRE*, 33, 709, October, 1945.

SEPARATION OF THE SYNCHRONIZING SIGNALS

51. Amplitude Separation. The separation of the composite synchronizing signal from the camera signal is performed after the composite video signal has been devel-

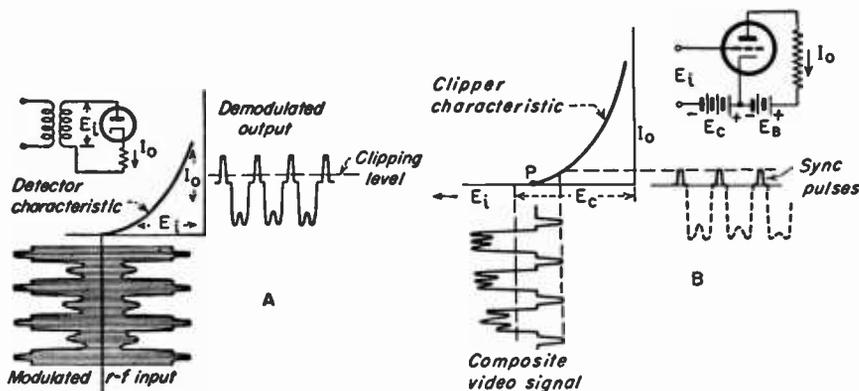


FIG. 48A. Demodulation of the modulated picture carrier by a diode detector, showing clipping level which separates the camera signal from the synchronizing pulses.

FIG. 48B. Clipper circuit and characteristic, used to separate composite sync signals from camera signals.

oped by the second detector. The composite video signal (Fig. 48B) is applied to a "clipper" tube, which is a tube that cuts off all current beyond a certain negative amplitude limit. A triode clipper tube and characteristic are shown in Fig. 48B. In

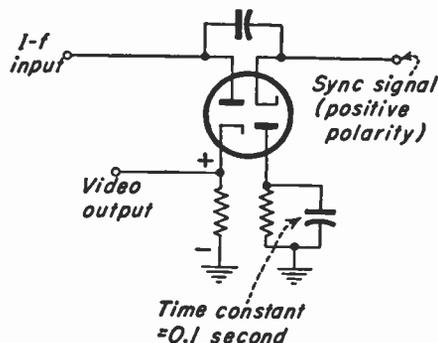


FIG. 49. Combined second detector and sync amplitude separator (clipper).

of pulses cannot be distinguished by amplitude means. Essentially wave-form separation depends on circuits which respond to the relative frequency content of the two

Fig. 49 a diode clipper arrangement is shown in conjunction with the second detector. It is necessary, of course, that the clipping level be maintained continuously at the blanking level to ensure that the camera signal does not affect synchronization, on the one hand, and to ensure that the maximum amplitude of sync pulses is developed, on the other.

52. Wave-form Separation. After the composite synchronizing signal has been separated from the video signal, it is necessary to develop the horizontal sync pulses independently of the vertical sync pulses. The latter separation is carried out by a method known as wave-form separation, since the two sets

of pulses cannot be distinguished by amplitude means. Essentially wave-form separation depends on circuits which respond to the relative frequency content of the two

sets of pulses. The horizontal sync pulses that are of short duration occur 15,750 times per second and have a predominance of h-f components, whereas the vertical pulses that are of long duration and occur 60 times per second have a predominance of l-f components. The ratio of the frequencies of the two sets of pulses $15,750/60 = 262\frac{1}{2}$ is the index of the degree of frequency difference on which the separator circuits may operate.

53. Differentiator Circuit for Horizontal Sync Pulses. The differentiator circuit shown in Fig. 50 is used to develop the h-f components of the composite synchronizing signal, i.e., the horizontal sync pulses. The series capacitance passes the high frequencies associated with the leading edge of the sync pulse, while retarding all lower frequency components. The RC product (time constant) of the combination is made short compared with the frame-repetition interval ($\frac{1}{60}$ sec) and long compared with the line scanning interval ($1/15,750$ sec). The leading edge of the differentiated wave forms is applied, in the proper polarity, to the synchronizing terminal of the horizontal scanning generator.

54. Integrator Circuit for Vertical Sync Pulses. The integrator circuit shown in Fig. 51 develops a sync pulse from the serrated vertical pulse and equalizing pulses. The wave forms of input and output are shown.

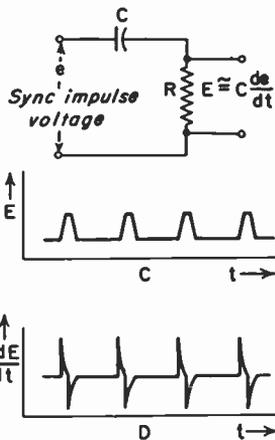


FIG. 50. (Top) Differentiator circuit. (Bottom) Action on sync pulses.

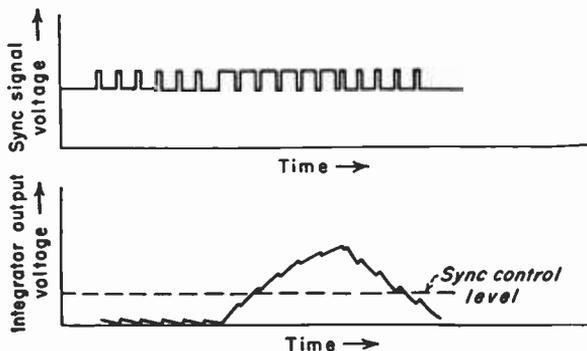
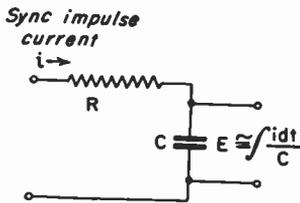


FIG. 51. (Top) Integrator circuit. (Bottom) Action on vertical sync pulse.

It will be noted that the initial portion of the integrated output pulse is not so sharply rising as that of the differentiated horizontal pulse, and consequently the intersection with the synchronizing control level is not so precisely marked. This fact makes it necessary to have the wave shape of each successive integrated pulse precisely the same. The function of the equalizing in this respect has been pointed out in Sec. 23. In addition it is necessary that all traces of the horizontal sync pulses be completely removed from the integrating circuit.

Several differentiating and integrating circuits may be used in cascade to improve the degree of separation. The cascaded circuits may be connected directly together (usually done with integrator circuits), or they may occur in the grid and plate

together (usually done with integrator circuits), or they may occur in the grid and plate

circuits of a sync separator amplifier tube. A typical synchronizing amplifier circuit is shown in Fig. 52.

PICTURE TUBES AND ASSOCIATED CIRCUITS

55. Picture Tubes. The conventional cathode-ray picture tube is a funnel-shaped evacuated structure containing an electron gun which forms an electron beam, and a

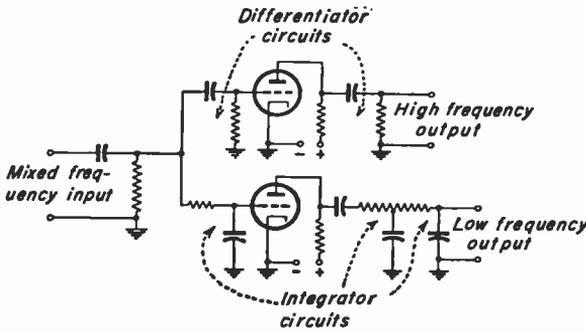


FIG. 52. Combined differentiator and integrator circuits for wave-form separation.

the cathode and the control electrode of the electron gun.

The beam is deflected synchronously with the scanning agent in the camera tube, and the beam current is controlled by the camera signal. The variations in the beam current produce corresponding variations in the brightness of the fluorescent spot, and the picture is thereby reproduced.

The operating characteristics of picture tubes depend on the design of the electron gun and on the physical and chemical properties of the fluorescent screen. The electron gun requires a power supply to form the electron beam. Finally the deflection fields must be provided by scanning generators, and these generators must operate under the control of the synchronizing signals of the video signal.

Picture tubes are classified according to (1) the type of focusing employed (electrostatic or magnetostatic) in the electron gun, (2) the type of deflection (electric or magnetic), (3) the type of phosphor (sulfide and nonsulfide), and (4) the color of the light produced (usually white).

56. Characteristics of Electron Guns. Electrostatically focused electron guns are characterized (1) by the ratio of the voltages applied to the second anode and the first anode. In present designs this ratio varies from 4 to 6. In addition the guns are characterized (2) by the control electrode characteristic which specifies the relation between control electrode voltage and beam current (second-anode current) for different values of second-anode voltage. A typical control characteristic of an electrostatically focused gun employed in the 12AP4 tube is shown in Fig. 54. Curves of this shape are typical of all types of electron guns, whether electrostatically or magnetostatically focused.

An important characteristic of electron gun is the degree of fineness of focus, *i.e.*, the size of the fluorescent spot formed on the screen. Guns of good design are capable of forming a fluorescent spot about 0.005 in. in diameter, but production tubes usually

fluorescent screen on which the beam impinges. The beam is deflected by the application of transverse electric or magnetic fields which cause the end of the beam to trace out the interlaced scanning pattern over the fluorescent screen. The current in the beam is capable of variation from zero (cutoff) to a maximum of several hundred microamperes, under the control of the signal potential applied between

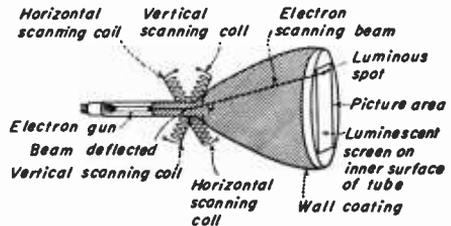


FIG. 53. Structure of typical picture tube.

have spots from 0.01 to 0.015 in. in diameter. The latter spot size permits a picture resolution of 350 lines when the picture height is 6 in. or more (picture width 8 in. or more).

57. Characteristics of Phosphors. The important operating characteristic of the phosphors (fluorescent materials) employed in picture tubes is the relationship between the light produced, the beam current (second-anode current), and the second-anode potential. Figure 55 shows a typical family of such curves, taken for the "P4" white-light phosphor. A recent improvement is the aluminum-backed phosphor, which has superior brightness and contrast when operated at voltages of 9,000 volts or higher.

58. Transfer Characteristic of Picture Tube. The transfer characteristic of a transducer in a television system is the relationship between the significant variational input quantity and the significant variational output quantity. In picture tubes the significant input is the control-electrode voltage, and the significant output is the corresponding light produced on the screen. The transfer characteristic of the 12AP4 tube is shown in Fig. 56. Note that the relationship is not linear but has the "antisaturation" shape. This characteristic tends to enhance the apparent contrast of the picture (see Sec. 72).

59. Contrast in Picture Tubes. The ratio of the brightness of the brightest portion of an image to the brightness of the darkest portion is called the *brightness-contrast range*. Owing to the effects of light spreading (halation) within the glass envelope of the tube the maximum contrast ratio of current picture tubes is about 100:1. Between closely adjacent portions of the image, halation reduces the maximum obtainable contrast to about 30:1.

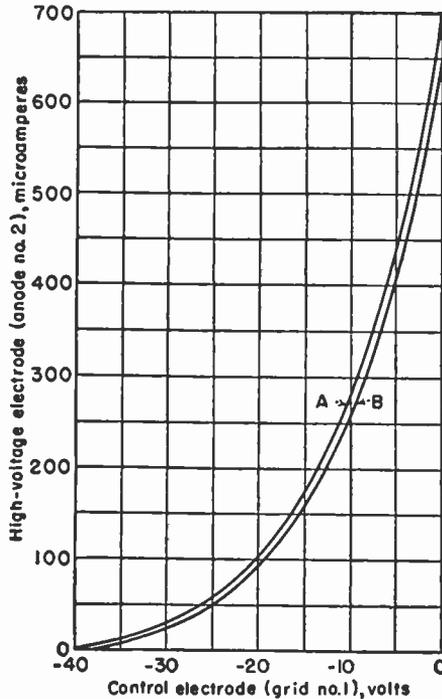


FIG. 54. Electron-gun control characteristic: (A) 7,000-volt, (B) 6,000-volt second-anode voltage.

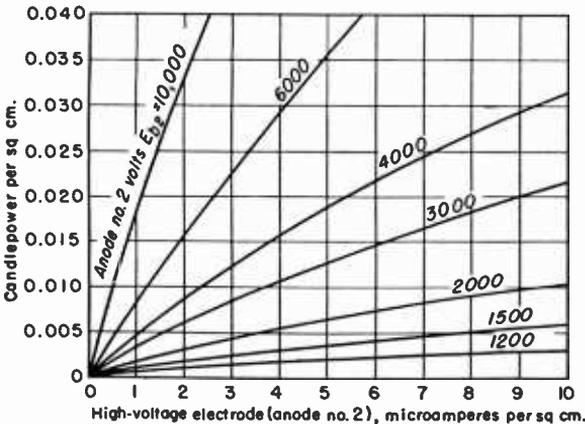


FIG. 55. Type P4 phosphor light-output characteristic.

applying the video signal wave form to the transfer characteristic (Fig. 57). The video

of an image to the brightness of the darkest portion is called the *brightness-contrast range*. Owing to the effects of light spreading (halation) within the glass envelope of the tube the maximum contrast ratio of current picture tubes is about 100:1. Between closely adjacent portions of the image, halation reduces the maximum obtainable contrast to about 30:1.

60. Dynamic Action of Picture-tube Control Circuit. The dynamic action of the picture-tube circuit is represented by

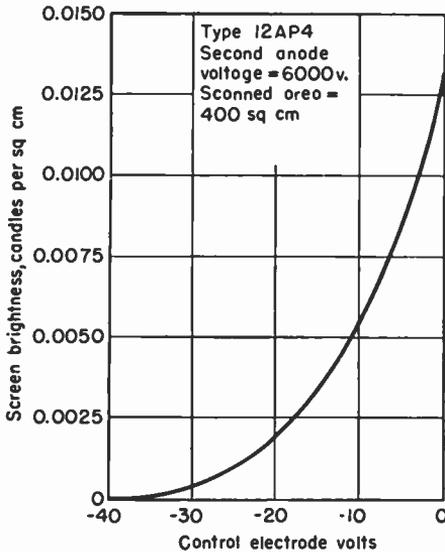


FIG. 56. Typical transfer characteristic of a picture tube, derived from Figs. 54 and 55.

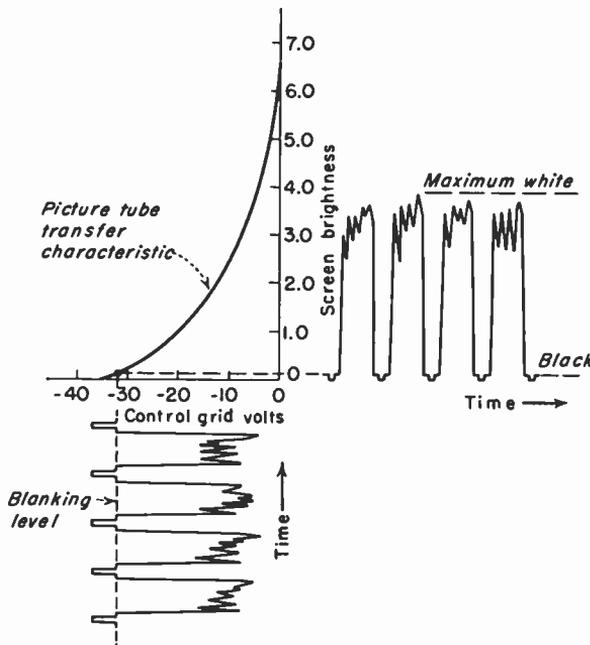


FIG. 57. Dynamic action of video signal in picture-tube control circuit.

wave form is applied so that the blanking level corresponds to the zero light (cutoff) point on the transfer characteristic as shown. This bias level must remain fixed at all times. Then the camera signal extending to the right of the blanking level produces light on the screen in accordance with the camera signal, whereas the synchronizing signals to the left of the blanking level are in the infrablack region (beyond cutoff) and do not produce light. The total excursion of the camera signal should be limited so that the control-electrode voltage never becomes positive; usually the control electrode does not go beyond the -5 or -10 volts mark. The average of the picture signal component, taken over the frame interval, establishes the background brightness of the scene, provided the blanking level remains fixed at the light cutoff point.

61. D-c Restoration Circuits. Two typical circuits used to maintain the blanking level constant at the picture-tube control electrode are shown in Fig. 58. The proportions of C_c and R_c are chosen to have a time constant long compared with the line-scanning interval but short compared to the duration of the changes in background light. In the upper diagram the grid and cathode of the video-amplifier tube act as a diode, whereas in the lower case a separate diode is employed. The diode, acting in conjunction with $C_c R_c$, develops a bias equal to the peak of the video signal. This peak value remains fixed (assuming no change in signal strength), consequently the remainder of the video signal (the camera-signal components) act in series with a fixed bias. This fixed

bias is coupled conductively (either through the amplifier tube, at the top, or directly, at the bottom) and forms a part of the control-electrode bias. By this means the blanking level remains fixed, and, if the total control-electrode bias is fixed so that the blanking level coincides with the light cutoff point, the background brightness of the scene depends only on the average of the camera-signal component, as is required.

62. Picture-tube Power Supplies.

The picture-tube power supply consists of (1) a source of high voltage for the first and second anodes, which act to draw the electrons from the gun and (in the case of electrostatically focused tubes) bring the beam to focus, (2) a source of heater current for the cathode of the electron gun, and (3) a source of focusing coil current (in the case of magnetostatically focused tubes).

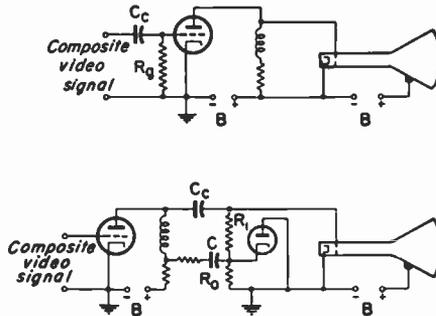


FIG. 58. Direct-current restoration circuits. (Top) Using the video amplifier grid current for rectification. (Bottom) Using a separate diode.

A typical high-voltage power supply is shown in Fig. 59. It consists of a single-winding transformer of rms output voltage equal to approximately $V_{dc}/1.4$, where V_{dc} is the desired output d-c voltage; two capacitors of roughly 0.03 to 0.05 μf ; a series filter resistor of roughly 100,000 to 500,000 ohms; and a tapped bleeder resistor of about 5 megohms. A resistor is also connected in series with the second-anode output tap to limit the total output current to a safe value in case of accidental contact by the operator. The taps required for the various electrodes of an electrostatically focused electron gun are shown.

The current required for the focusing coil of a magnetostatically focused gun depends on the focus-coil design. A typical value is 100 ma at 25 volts, which may be obtained from

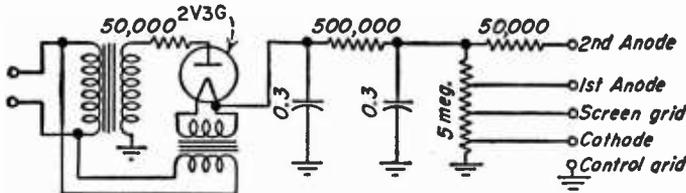


FIG. 59. Typical anode voltage supply for a picture tube.

the current drain of the receiver proper at the sacrifice of 25 volts in the low-voltage power supply for the receiver.

The heater currents have usually one of two rms a-c values: 2.5 volts at 2.1 amp, or 6.3 volts at 0.6 amp.

The high-voltage a-c power may also be derived from (1) high-voltage pulses appearing across the horizontal scanning transformer primary or (2) a h-f oscillator. In the latter case the filtering is simplified, since small values of filter capacitance suffice.

DEFLECTION OF ELECTRON BEAMS

63. Electron-beam Velocity. The amount of deflection suffered by an electron scanning beam depends on the velocity with which the electrons in the beam move. This velocity v is expressed by

$$v = 3 \times 10^{10} \sqrt{1 - \left(\frac{1}{2 \times 10^{-6} E + 1} \right)^2} \quad \text{cm per sec} \quad (32)$$

where E is the accelerating voltage in volts (approximately equal to the second-anode voltage). This expression takes into account the change in electron mass with veloc-

ity. The values of v vary from 2.66×10^9 cm per sec at 2,000 volts to 9.86×10^9 cm per sec at 30,000 volts.

64. Electric Deflection. The deflection d , in centimeters, of the scanning beam across the screen of a picture tube caused by passage between parallel deflecting plates is given by

$$d = \frac{1.77 \times 10^{15} E_d l (D + \frac{1}{2}l)}{sv^2} \quad \text{cm} \quad (33)$$

where E_d = voltage applied to deflection plates, volts

v = electron beam velocity, cm per sec

l = length of deflection plates, cm

s = separation between them, cm

D = distance from screen end of deflection plates to center of screen measured along axis of tube, cm

Typical electrically deflected tubes have deflection sensitivities of from 0.15 to 0.35 mm deflection per volt applied to the deflecting plates, when operated at maximum rated second-anode voltage.

65. Magnetic Deflection. The deflection d , in centimeters, across the screen of a picture tube, caused by passage through a uniform magnetic field is given by

$$d = \frac{1.77 \times 10^7 B l D}{v} \quad \text{cm} \quad (34)$$

where B = flux density of field, gauss

l = its length, cm

D = field-to-screen distance, cm

v = electron-beam velocity, cm per sec

66. Ion Spot. Negative ions liberated from the cathode of the electron gun are focused and deflected in much the same manner as the electrons. In electric deflection the deflection is independent of the charge/mass ratio of the particles; hence the ions and electrons are equally deflected. In magnetic deflection, however, the deflection depends on the square root of the charge/mass ratio. Since the ions have masses several thousand times that of the electron, they suffer correspondingly small deflection. The lack of deflection subjects the center of the scanned area to continual bombardment by the ions, and this eventually results in the formation of a black or yellowish spot. The ion spot is characteristic of the combination of electrostatic focusing and magnetic deflection. Specialized electrode structures have been devised, however, which intercept the ions before they reach the screen. The aluminum-backed phosphor (Sec. 57) is free from ion spot since the aluminum inhibits ionic bombardment.

SCANNING AND SYNCHRONIZATION

67. Saw-tooth Generators. The saw-tooth wave form (Fig. 7) is generated for scanning purposes by the periodic charging and discharging of a capacitor. The charge-time curve is used to produce the active scanning motion, and the discharge curve forms the retrace. To maintain a linear charge curve, it is customary to restrict the charge time to about 0.4 time the RC product of the circuit, or less, and also to make use of the nonlinear dynamic characteristic of the following amplifier to introduce a compensating nonlinearity. Certain forms of multivibrator circuits may be used to produce saw-tooth waves directly.

Usually a separate *discharge tube* is used to discharge the capacitor. The discharge current is passed through a high-vacuum triode whose grid controls the timing of the discharge. The impulses applied to the grid of the discharge tube are usually derived from an impulse generator, although they may consist of the synchronizing signal itself, properly amplified.

Impulse generators used to control the discharge tube in scanning generators take one of two forms, the multivibrator or the blocking oscillator. The blocking oscillator

(Fig. 60) consists of a grid-plate-coupled oscillator whose grid is driven negative by the passage of grid current, thus blocking the oscillations suddenly. As the charge leaks off the grid through the grid resistor, the oscillations recommence, to be followed by the sudden blocking of the grid circuit. The sharp impulses appearing between the grid and ground are used to control the discharge tube as shown.

A multivibrator type of saw-tooth generator is shown in Fig. 61. This circuit operates by virtue of the connection between the plate circuit of the output tube and the grid circuit of the input tube. The alternate charge and discharge of the coupling

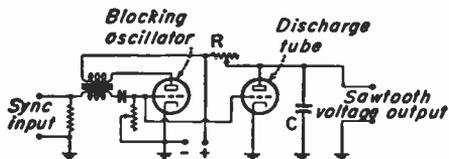


FIG. 60. Blocking-oscillator type of impulse generator.

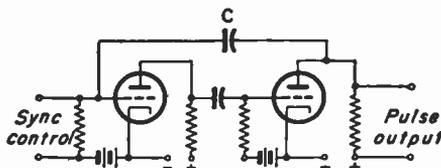


FIG. 61. Multivibrator type of impulse generator.

capacitor can be used to produce either impulses or saw-tooth waves, depending on the circuit constants.

68. Production of Current Saw-tooth Waves. Saw-tooth waves of voltage produced by saw-tooth generators suffice to deflect the beam of an electrically deflected tube, which is a voltage-operated device, provided only that the peak-to-peak value of the saw-tooth wave is great enough to produce full deflection. In magnetically deflected tubes the deflection is proportional to the current in the deflection coils; hence saw-tooth waves of current are required. The voltage wave form required to produce saw-tooth waves of current depends on the inductance and resistance present

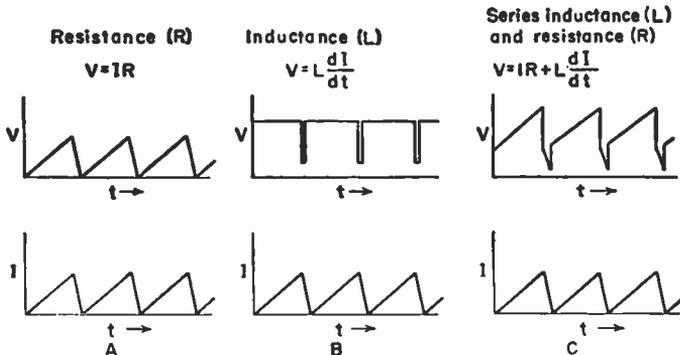


FIG. 62. Saw-tooth waves of current and corresponding voltage wave forms in circuits of R, L, and LR.

in the scanning-coil windings. An "impulse" voltage wave is required for coils exhibiting a large inductance-resistance ratio. For lower L/R ratios the voltage wave form is a combination of impulse and saw-tooth waves. The several voltage and current wave forms for these cases are shown in Fig. 62.

The part-impulse part-saw-tooth wave form may be produced simply by applying a saw-tooth wave to a series RC combination. The saw-tooth component develops across the resistance, while the impulse portion develops across the capacitance.

The presence of distributed capacitance in the scanning-coil windings gives rise to resonance oscillations when the impulse voltage wave form is applied. These residual

oscillations may be damped out by connecting a rectifier tube and a shunt RC circuit in series across the scanning-coil terminals.

69. Amplification of Scanning Wave Forms. The preservation of the scanning wave form in the amplifier subsequent to the scanning generator is based on the considerations for video amplifiers. Usually it is desirable to pass the fundamental and 20 harmonics, which makes the range 60 to 1,200 cps for the vertical scanning amplifier and 15,750 to 315,000 cps for the horizontal amplifier. The phase and amplitude characteristics must be linear over these ranges.

For electric deflection it is essential that the scanning-generator output be disposed symmetrically with respect to the deflection plates, and this is carried out by employing push-pull amplification. The center point of the push-pull output is connected through a high resistance to the second-anode terminal of the picture tube. Care must be taken to allow the necessary peak-to-peak voltage to develop across the amplifier output without breakdown of insulation and excessive stress in the tube structures. The necessity for high scanning voltages has limited application of electric deflection to tubes operating below 6,000 volts, second-anode voltage.

In magnetic deflection, heavy current rather than high voltage is required to secure full deflection. To secure the current, it is customary to employ a voltage step-down transformer in the output of the scanning amplifier. This transformer must meet the amplitude- and phase-frequency characteristics of the amplifier itself. High voltage

develops across the primary of this transformer as a result of the rapid changes of current in the secondary. The amplifier tubes and other components must be capable of withstanding these voltage peaks, which often attain several thousand volts amplitude.

70. Scanning Yokes.

The set of deflection coils required for magnetic deflection is called a *scanning yoke*. It consists of two sets of coils. One, arranged about a vertical axis transverse to the

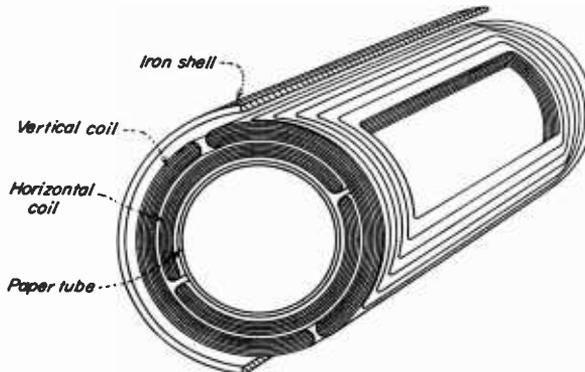


FIG. 63. Construction of a typical scanning yoke (magnetic-deflection coil system).

tube axis, produces the horizontal deflection, and another set of coils, arranged on a horizontal axis transversely to the tube axis, produces the vertical deflection.

Among the factors on which the yoke design depends are (1) the angle of deflection required, which determines the required number of ampere-turns as well as the allowable physical length of the yoke; (2) the necessity of providing a uniform field, to avoid defocusing the spot and distorting the orthogonal shape of the scanning pattern; and (3) the proportioning of the L/R ratio to secure linear deflection with a given deflection amplifier and output transformer.

CONTRAST AND GRADATION OF TELEVISION IMAGES

71. Over-all Brightness Transfer Characteristic. The ability of the television system to reproduce brightness contrasts and tonal gradations is expressed by the over-all brightness transfer characteristic (Fig. 64). The ordinates give the range of brightness in the reproduced image (image brightness) corresponding to the range of brightness in the original object (object brightness) plotted in the abscissas.

The actual shape of this curve depends on the transfer characteristic (input-output relationship) of each item of equipment in the transmission system. In general the actual

characteristics cannot be expressed in simple analytic form. However, if an idealization is made, the curves may be expressed in the following form:

$$B_i = B_o k_o \gamma_o \tag{35}$$

where B_i is the image brightness corresponding to the object brightness B_o , k_o is the proportionality factor relating the image brightness scale to the object brightness scale, and the exponent γ_o ("transfer slope") determines the extent and direction of the curvature of the characteristics. For unity slope ($\gamma_o = 1$) the relationship between B_i and B_o is linear. For slope greater than unity, the curve has an "antisaturation" shape; for slope values below unity, the curve has a "saturation" shape.

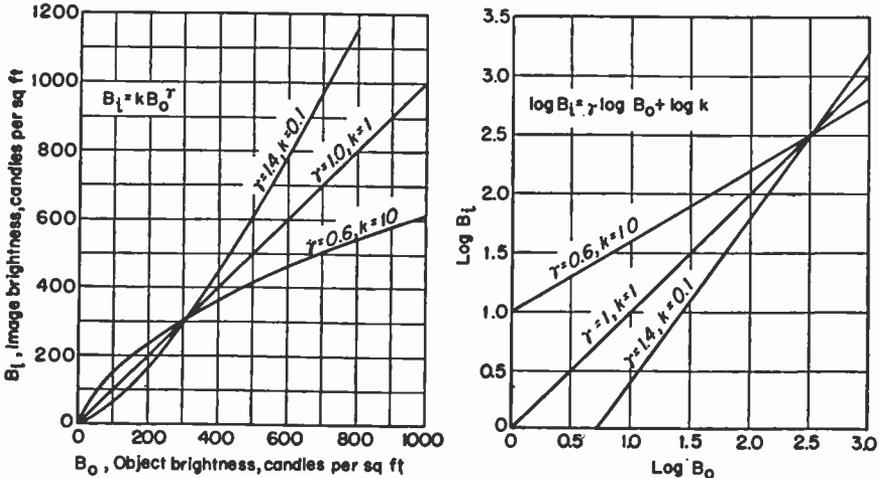


FIG. 64. Brightness transfer characteristics of a television system.

The value of slope determines the subjective contrast of the image as viewed by the observer, since the sensation of light in the mind is approximately proportional to the logarithm of the brightness. When Eq. (35) is expressed in logarithmic form

$$\log B_i = \log k_o + \gamma_o \log B_o \tag{36}$$

all the relationships between $\log B_i$ and $\log B_o$ become linear and the slope of the lines is directly proportional to the slope value of γ_o . In consequence high contrast is produced by correspondingly high values of slope.

72. Subsidiary Transfer Characteristics. The input-output characteristics of each subsidiary item of equipment in the system can be expressed by a similar relationship

$$\text{Output} = k (\text{input})^\gamma$$

where k relates the scales of the input and output quantities and γ is the slope exponent describing the curvature of the characteristics. By combining each curve in the transmission system successively, equating the output of one device to the input of the succeeding device, it can be shown that the over-all slope of the system is equal to the product of all the subsidiary slopes. In consequence, the effect of one item of equipment whose slope is lower than unity may be compensated by that of another whose slope is the inverse of the first. The slope of iconoscope tubes, for example, lies at about 0.7, whereas that of picture tubes is about 2.5. Assuming that the subsidiary amplifiers, modulators, and demodulators are linear (slope unity), the over-all slope is then $0.7 \times 2.5 = 1.8$; *i.e.*, the slope is somewhat above unity. The orthicon camera, on the other hand, has a slope of unity, and the over-all slope in this case would be 2.5, producing a considerably more contrasty reproduced image. The desirable value of over-all slope, following motion-picture practice, is between 1.2 and 1.7. The high value of slope aids in restoring color contrasts

lost through the monochromatic nature of the reproduction. It should be noted that high contrasts are limited by picture-tube performance.

The values of the subsidiary slopes also bear on the signal/noise ratio of the system. If a transmitter slope less than unity is employed, most of the picture information consists of signal excursions having amplitudes high on the dynamic characteristic,

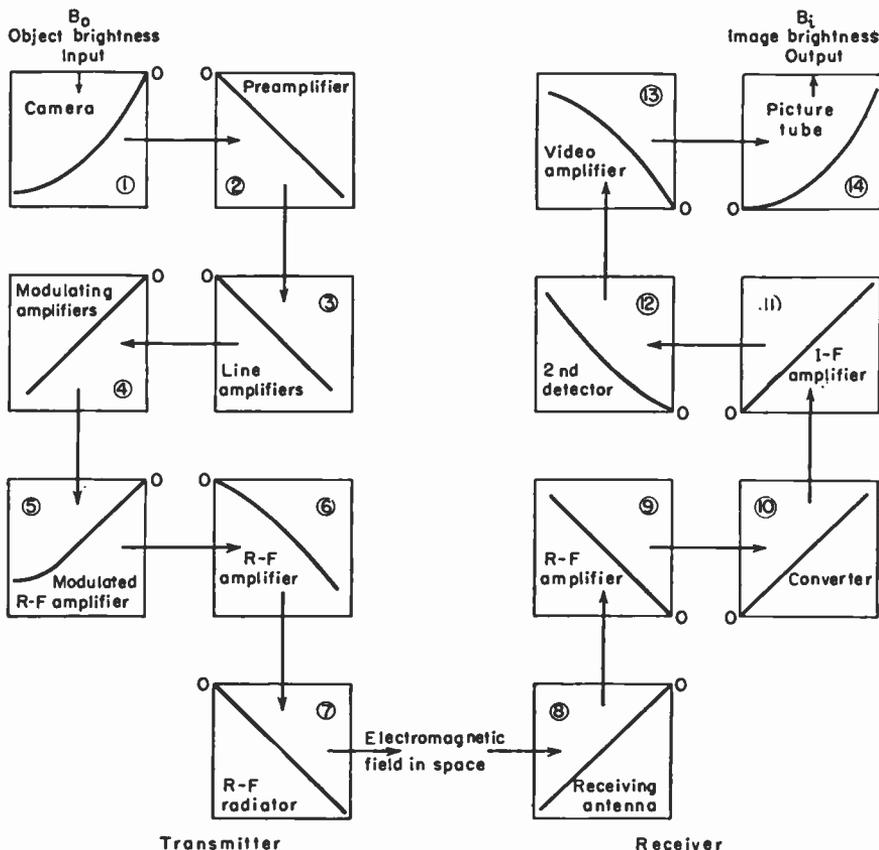


FIG. 65. Transfer of object brightness to image brightness through subsidiary transfer characteristics of the elements of a television system.

above the noise. A compensating higher value of slope in the receiver may be used to produce an over-all value within the desirable range of 1.2 to 1.7.

References

Picture Tubes and Image Reproduction:

- Bachman and Carnahan: Negative Ion Components in the Cathode Ray Beam, *Proc. IRE*, **26**, 529, May, 1938.
- Burnett, C. E.: A Circuit for Studying Kinescope Resolution, *Proc. IRE*, **24**, 1095, August, 1937.
- Epstein, D. W.: Electron Optical System of Two Cylinders as Applied to Cathode-ray Tubes, *Proc. IRE*, **24**, 1095, August, 1936.
- Iams, H.: A Fixed Focus Electron Gun for Cathode-ray Tubes, *Proc. IRE*, **27**, 103, February, 1939.
- Langmuir, D. B.: Theoretical Limitations of Cathode-ray Tubes, *Proc. IRE*, **25**, 977, August, 1937.
- Law, R. R.: Contrast in Kinescopes, *Proc. IRE*, **27**, 496, July, 1939.

Leverenz, H. W.: Problems Concerning the Production of Cathode-ray Screens, *J. Opt. Soc. Am.*, **27**, 25, January, 1937.
 Levy and West: Luminescence and Its Application to Television, *J. Telev. Soc.*, **2**, 337, March, 1938
 Maloff, I. G.: Gamma and Range in Television, *RCA Rev.*, **3** (4), 409, April, 1939.
 ——— and D. W. Epstein: Theory of the Electron Gun, *Proc. IRE*, **22**, 1386, December, 1934.
 ——— and ———: "Electron Optics in Television," Chap. XII, McGraw-Hill, 1938.
 Zworykin and Painter: Development of the Projection Kinescope, *Proc. IRE*, **25**, 937, August, 1937.

COLOR TELEVISION

73. Development of Color Television. Since 1940 several groups of workers in the United States have been developing color television for public service. The *sequential system*, developed by P. C. Goldmark and his associates, transmits images in three

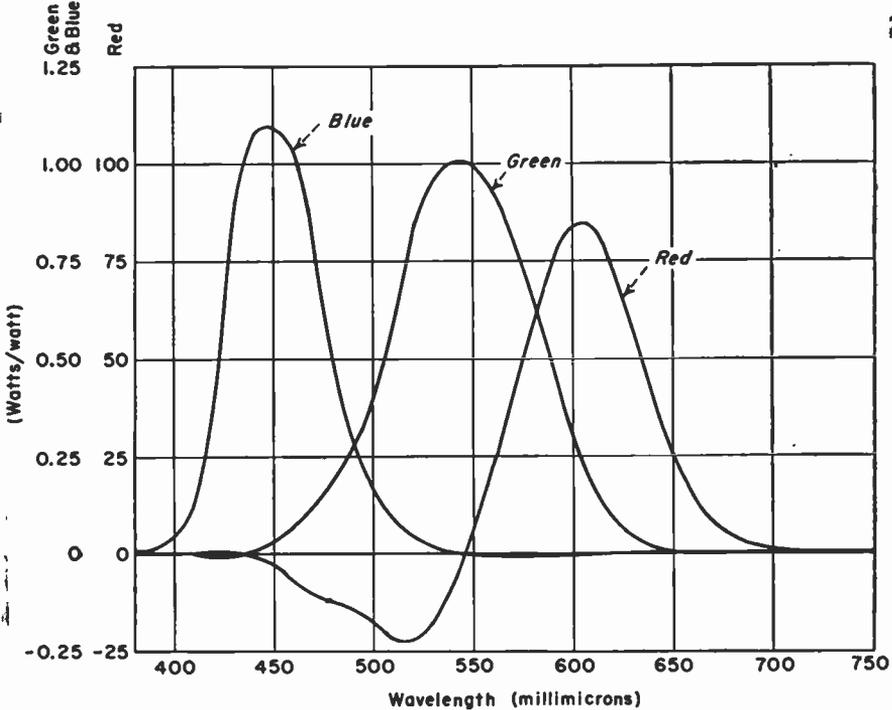


FIG. 66. Relative amounts of standard primary colors required to match spectral colors, as per ICI "standard observer."

primary colors one after another over a single channel. The *simultaneous system*, developed by R. D. Kell and his group, transmits images in three primary colors at the same time over three narrower channels. The sequential system employs simpler terminal equipment (one pickup device and picture tube), and uses the full band width for each color. The simultaneous system requires more elaborate terminal equipment (three pickup and reproducing devices), but permits less band width to be employed in the blue channel than in the green and red channels, thus offering a saving in the over-all band width. Moreover the simultaneous system, by suitable choice of the scanning standards, may be made compatible with the black-and-white system, permitting the color transmissions to be received in monochrome on receivers built for the present black-and-white standards. Both systems make use of the same principles of color synthesis and are capable of equal fidelity of color rendition.

74. Color Matching. In color television, the system must be able to achieve a match between colors in the original subject and in the televised reproduction. Color matching is greatly simplified by the fact that color perception in the human eye is based on three sets of color receptors in the central portion of the retina. These are sensitive over three contiguous portions of the visible spectrum, centering in the red, green, and blue regions. The spectral sensitivity of each group of receptors may be measured by noting the relative amounts of three spectral primary colors required to match the various colors of the spectrum. Figure 66 shows the relative amounts of standard red, green, and blue spectral primaries required to match each color in the spectrum, based on the "standard observer" of the International Commission on Illumination (ICI).

By so combining three primary colors it is possible to produce a very close match to any color in the spectrum, as well as colors (such as brown, purple, magenta, cyan) which do not occur in the spectrum.

The range of colors over which such matches can be made depends on the number and character of the primary colors used. Two primary colors suffice to match only a very narrow range of colors, while three properly chosen primaries suffice to match closely nearly all the colors that occur in nature. More than three primary colors serve no practical purpose.

A convenient graphical representation of the color-matching process is the color diagram (properly, "chromaticity diagram") shown in Fig. 67. This is a two-dimensional representation of three quantities, x , y , and z , known as the trichromatic coefficients. The quantities x and y are shown explicitly on the diagram, and z coordinate is given implicitly by the relation $x + y + z = 1$. These quantities specify in normalized fashion the relative proportions of three standard primaries required to match particular colors. A point on the diagram represents one such mixture of primaries and corresponds to the color matched by that mixture. If points are plotted, representing the mixtures required to match each color as it occurs in the spectrum, the points fall on a locus ("spectral locus") shown as a solid line in the figure. The blue colors occur near the origin of the coordinates, the green colors at the top, near the y axis, and the red to the right in the diagram. The spectral locus is closed by a straight line between its extremes. All colors visible to the ICI standard observer may be represented as points in the area bounded by this roughly triangular figure.

The convenience of the color triangle resides in the fact that a straight line, drawn between points representing two colors, contains points representative of all the colors which can be obtained by combining the two end-point colors in various proportions. Moreover, if three points are chosen and lines are drawn between them to form a triangle, the triangle so formed encloses points representative of all the colors that can be formed by mixing the three apex-point colors in various proportions. In particular, if lines are drawn from an interior point to the apex points, the lengths of these lines are in inverse ratio to the amounts of each apex-point color required in the mixture to match the color represented by the interior point.

The standard ICI spectral primaries (on which Fig. 66 is based) are shown in Fig. 67 as B_s (blue, 435.8 $m\mu$), G_s (green, 546.1 $m\mu$), and R_s (red, 700 $m\mu$). The triangle formed by these points contains all the colors that can be matched by simple addition of these primaries. Although a portion of the area within the spectral locus is not included in the triangle, the portion not included contains shades of green and blue which are not clearly distinguished from one another by the eye, and which in any event do not occur in nature to any extent. The colors included within the three primaries include the vast majority of distinguishable hues and saturations; omission of colors external to the triangle formed by the primaries does not preclude accurate reproduction of colors encountered in practice.

White light is represented in the central region of the chart. The point marked C represents a standard white (ICI "illuminant C ") which closely approximates the color of daylight from the northern sky plus sunlight. Any three colors surrounding this point will, when combined in the proper proportions, match this white. In fact

any two colors, the line between which passes through *C*, will, when properly combined, match this standard white. But such a two-primary system will not match colors off the line and so gives a very poor approximation to nature. For this reason the two-primary system, although occasionally used for motion-picture work, has been abandoned in color-television development.

In color photography and television it is not convenient to employ spectral primary colors, which can be obtained only by selecting a narrow portion of a spectrum produced by a prism or grating. Rather, primary colors are produced by passing light

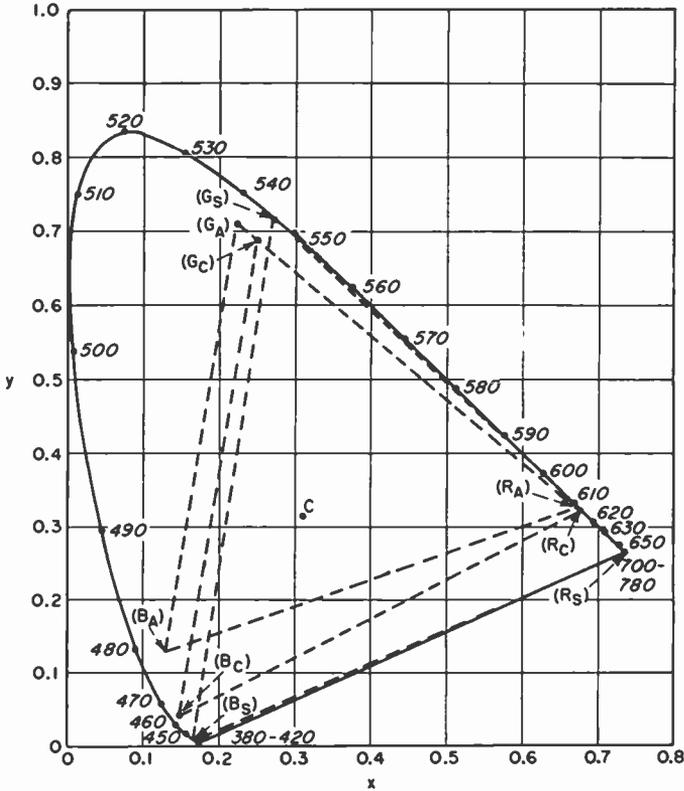


FIG. 67. Chromaticity diagram, showing color gamuts covered by various primaries suggested for color television.

through filters, *i.e.*, transparent pieces of colored gelatin, glass, or plastic. These filters are interposed between the subject and the camera, and between the picture tube screen and the audience. Standard filters of this type, such as the Wratten series, are widely available.

In a television system, the primary colors used at the transmitter are compounded of three characteristics: the spectral output of the light source, the spectral transmission characteristic of the filter material, and the spectral sensitivity of the photosensitive surface in the camera. At the receiver the corresponding primaries are compounded of the spectral transmission of the filter, and the spectral output of the cathode-ray phosphor. Typical over-all television receiver primaries, used by Goldmark in his sequential system, are identified on the color diagram (Fig. 67) as *B_s*, *G_s*,

R_c . The area contained within the triangle passing through these points is seen to be somewhat smaller than that contained by the spectral primaries (B_c, G_c, R_c), but the range of colors reproducible with the television primaries is sufficient for a very high degree of realism. The area contained by the television primaries is greater than that covered by the photographic transparency processes (Technicolor, Kodachrome,

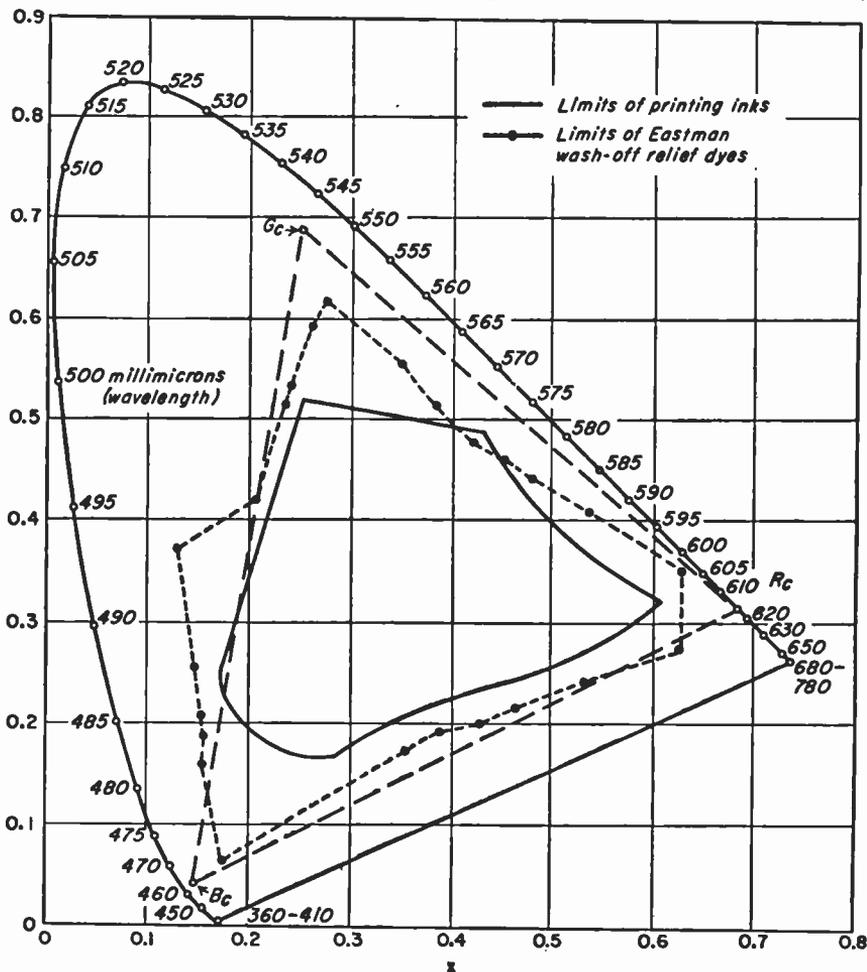


FIG. 68. Chromaticity diagram, showing color gamuts of printing inks and photographic dyes.

and AnscoColor). The primaries marked $B_c, G_c,$ and R_c are those recommended for the 180-field sequential system and the 60-field simultaneous system.

Figure 68 shows the area of the color diagram occupied by other methods of color reproduction, *i.e.*, printing inks and wash-off relief dyes used in photographic color prints. It will be noted that the television primaries offer a wider gamut of colors than do these widely accepted methods of color reproduction.

The choice of primaries for a television receiver involves also the question of flicker.

In the sequential system, if the average light transmission (luminosity) of one primary filter is substantially different from that of the other two, the flicker will appear at lower brightness levels than if all three filters have more nearly the same light-transmitting ability. In the simultaneous system this effect does not occur.

75. Color Balance. In color reproduction it is necessary that the color match between original subject and televised reproduction be correct over the range of brightness present in the original scene, and moreover that the colors appear relatively correct when the brightness range of the reproduction is substantially less than that of the original scene. It is thus fortunate that the color-matching process represented in the color diagram applies over a wide range of absolute brightnesses. Thus, suppose a given blue is matched by a particular mixture of primaries. If the brightness of the blue is increased substantially so that it appears to be blue-white, the same mixture of primaries will match the blue-white, provided only that the brightness of each primary is increased in the same proportion.

To achieve color balance over a range of brightness, it is necessary that the intensity of the primaries shall vary in strict proportion to the change in the brightness of the original color. Thus if the transfer characteristic (see Sec. 71) that applies to the green primary does not have the same extent and shape as those applying to the red and blue, correct color matching will be achieved only over a restricted range of brightness. To avoid this restriction, it is desirable to employ dynamic control amplifiers

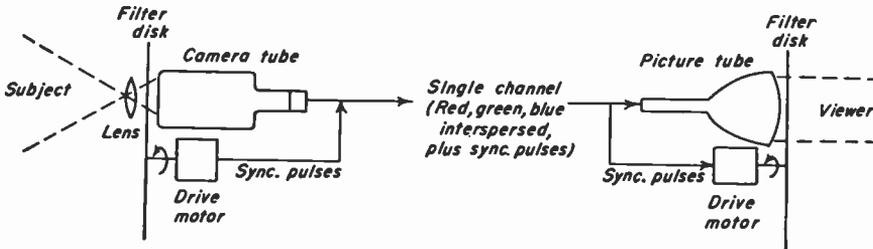


FIG. 69. Sequential color-television system.

which will permit the shape and extent of one transfer characteristic to be matched to the others. To permit accurate delineation of colors over a range of brightness, it is necessary that all three transfer characteristics be linear, *i.e.*, the log of each receiver primary brightness should be a linear function of the log of the corresponding transmitter primary brightness. But such accurate reproduction of colors may not produce the most pleasing effect, particularly if the range of brightness is compressed in the reproduction. Hence, as in other color-reproduction systems, it is desirable to introduce controlled distortions to the transfer characteristics of each primary. Transfer-characteristic control amplifiers, previously referred to, permit such distortions to be introduced.

A critical test of the ability of a color-reproduction system to maintain color balance is the transmission of a gray scale, *i.e.*, a series of gray patches having brightnesses in logarithmic progression. Lack of color balance is revealed by the presence of a tinge of color in one or more of the patches of the reproduced scale. Adjustment of the transfer-characteristic control amplifier, to remove any color tinge from the steps of the gray scale, will ensure accurate (but, as previously noted, not necessarily the most pleasing) reproduction of color values.

76. Sequential Color Systems. In the sequential system, (Fig. 69) a single camera tube (which may be an iconoscope, orthicon, or image dissector) views the subject through a rotating disk carrying segments of red, blue, and green filters. One complete field (set of alternate lines) is scanned while a filter of one color lies between the subject and the camera. As successive fields are scanned, successively different colored filter segments rotate into position. The shape of the filter segments is so pro-

portioned that, as the disk rotates at constant speed, light of one primary color enters the camera only during the scanning of each separate field.

A point on the scanning pattern is scanned in all three colors only after six complete fields have been scanned, in the following sequence:

Frame No.	Field No.	Primary Color
1.....	1 (odd)	Red
	2 (even)	Blue
2.....	3 (odd)	Green
	4 (even)	Red
3.....	5 (odd)	Blue
	6 (even)	Green

If a portion of the televised subject has a color closely approximating one of the primary colors, that portion is reproduced at a rate which is effectively one-third the field rate of the scanning pattern. Colors intermediate to the primary colors are reproduced at rates higher than one-third the field rate but not so high as the field rate. Flicker between lines (not visible at normal viewing distances) occurs at a rate one-sixth the field rate.

Thus, in the sequential system, the picture repetition rate is generally lower than that of a black-and-white system using the same field rate, so flicker appears in the color images at correspondingly lower values of brightness. To avoid flicker at viewing brightnesses comparable to those demanded in black-and-white service, it is necessary to increase the field rate of the sequential system to a value substantially higher than that suitable for the black-and-white system. Goldmark has employed a field rate of 144 per sec, compared with 60 per sec for black-and-white images. Other workers have urged that the field rate should be 180 per sec (three times the black-and-white rate) to ensure absence of flicker at brightness levels comparable to those achieved in the black-and-white system. Since the scanning rate is thus increased by a factor of 2.4 to 3 times, the highest v.f. produced in the scanning process is 2.4 to 3 times as great, assuming an equal number of lines in the image. On the basis of a 525-line image, the maximum v.f. required for the sequential color system is 9.6 to 12 Mc, depending on the frame rate used.

At the receiver in the sequential system, a corresponding filter disk is set up between the picture-tube screen (which has a specified white color) and the viewer. This disk is rotated in synchronism with the filter disk at the transmitter. Thus as the televised subject is scanned in a field of one color, the corresponding field appears at the receiver in the same color. As the colored fields follow each other in the sequence outlined above, the images fuse in the mind of the observer, so the blended value of color in the image is recreated.

By careful choice of the light source and spectral sensitivity of the camera tube at the transmitter, and the phosphor at the receiver, it is possible to select appropriate primary colors for the filter segments and achieve the coverage of the color diagram shown in Fig. 67. Consequently the colors reproduced in the sequential system are highly realistic.

The principal shortcoming of the sequential system is the tendency for flicker to appear, when the image has the brightness required for satisfaction in home use, unless a high field rate (180 per sec) is used. This in turn requires a high speed of rotation of the filter disks with accompanying mechanical problems. Since the frame rate is inevitably two to three times that required in black and white, it is not possible to receive high-definition sequential color images on black-and-white receivers without adjusting the scanning rates and broadening the pass band of the receiver circuits. The sequential color system is thus not compatible with the black-and-white system.

The sequential system need not employ rotating parts. If three camera tubes are used, each fitted with a separate fixed filter and focused in register on the subject, and

if the images from three corresponding picture tubes with fixed filters are projected in register at the receiver, the cameras and picture tubes may be keyed on and off synchronously, in sequence, to reproduce the image.

In addition to the flicker problem, the sequential system exhibits two effects which may detract from the quality of the reproduction: color breakup and color fringing. Color fringing occurs when the motion of the subject relative to the field of view is so rapid that edges of the subject are scanned in one color only while it moves an appreciable distance. Such rapid motion is accompanied by the appearance of color fringes. Color breakup occurs when the eye moves rapidly while viewing a sequential image, in response to motion in the image or because of motions of the neck or eye muscles. When this occurs, the successive color scanings of the subject fall out of register on the retina. These effects are not so noticeable as to detract seriously from the enjoyment of sequential images, but they may represent a marginal lack of quality, compared to simultaneous images, in which neither effect occurs. In the simultaneous system, however, lack of color register may produce equally disadvantageous color fringes.

77. Standards for the Sequential System. An extensive study of standards for sequential transmissions was undertaken by the Radio Technical Planning Board's (RTPB) Panel on television to prepare evidence for a hearing held before the FCC from November, 1946, to February, 1947. Substantial agreement was obtained on the following items:

1. *Scanning Pattern.* 525 lines per frame, interlaced in two fields of 262.5 lines each, and displayed in an aspect ratio of 4:3.
2. *Color Sequence and Field Rate.* Successive fields to be scanned in the order red, green, blue at a rate of 144 or 180 color fields per second. The majority of the panel preferred the value of 180 fields per second, as offering the best performance with respect to flicker and brightness.
3. *Primary Colors.* The over-all primary colors (combined effect of phosphor and filter) of an ideal receiver were recommended as follows (transmitter primaries would be chosen to match the ideal receiver primaries):

Primary color	Trichromatic coefficients		Primary color	Trichromatic coefficients	
	<i>x</i>	<i>y</i>		<i>x</i>	<i>y</i>
Blue (<i>B_s</i>).....	0.130	0.1305	Blue (<i>B_t</i>).....	0.1477	0.0412
Red (<i>R_s</i>).....	0.675	0.326	Red (<i>R_t</i>).....	0.6805	0.3193
Green (<i>G_s</i>).....	0.221	0.712	Green (<i>G_t</i>).....	0.250	0.6885

The primary filters for the 144-field system have more nearly equal luminosities and hence less tendency to produce flicker. They do not enclose so large an area of the color diagram as the 180-field primaries, however, and hence do not reproduce so wide a gamut of color.

4. Reproduction of White. The ideal receiver should reproduce the standard white light, illuminant *C*, when equal signal amplitudes are applied on successive fields. Correspondingly, when the transmitter views a magnesium-oxide block illuminated with this standard white, it should produce equal signal amplitudes on successive fields.

5. R-f Channel. No recommendations concerning methods of modulation and channel width were made by the RTPB, although the general sentiment of the panel was for vestigial-side-band a.m. Typical channels based on this method of modulation, for 144- and 180-field systems, are shown in Fig. 70.

The opinion of the RTPB panel was that no standards should be set up for commercial operation of a color-television system pending further exploration of the relative advantages of the sequential and simultaneous systems.

Consideration has been given to a 405-line, 144-frame sequential color system, operating on a 6-Mc channel, at a hearing before the FCC in September, 1949.

78. Simultaneous Color System. The simultaneous system (Fig. 71) transmits picture information in the three primary colors simultaneously, using separate channels. In essence the system comprises three black-and-white systems, each transmitting pictures in one primary color only, the cameras and picture tubes being so arranged that the images are picked up and reproduced in optical register. The terminal and modulation equipment of the system is evidently more complex than that of the rotating-disk sequential system, but this disadvantage is offset by several important advantages. Since the simultaneous system uses three light sources at the receiver, for a given picture size, phosphor efficiency, and accelerating voltage, the

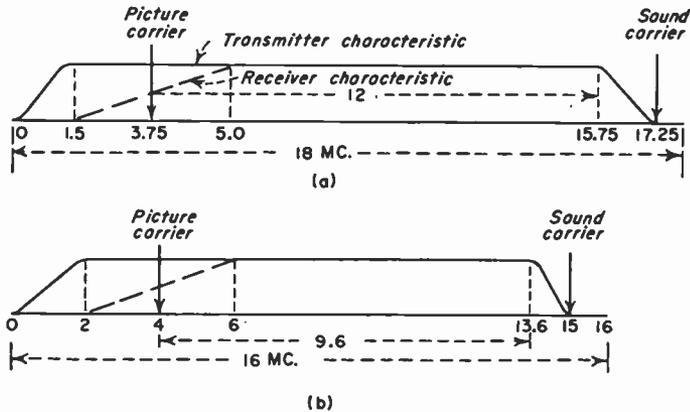


FIG. 70. Arrangement of channels for a 525-line sequential color system. (Top) 180 frames per second. (Bottom) 144 frames per second.

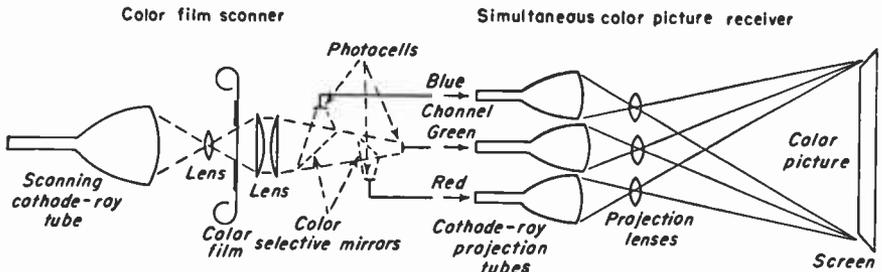


FIG. 71. Simultaneous color system.

simultaneous reproduction is approximately three times as bright as the equivalent sequential image. Moreover, this increased brightness does not bring with it a flicker problem, since each color is present simultaneously, rather than interspersed in time. In fact, the flicker performance of the simultaneous system is the same as that of a black-and-white system using the same frame and field rates. The color fringing and breakup phenomena are not present. Finally, the detail transmitted in each color may be chosen to meet the requirements of the eye, and substantial saving in the overall band width of the system thereby achieved.

The principal difficulty in the terminal apparatus of the simultaneous system is in securing accurate optical register of the superimposed primary images. At the transmitter a single scanning agent may be used, in the form of a flying spot scanner (Fig. 71) which excites three separate phototubes viewing the scene through fixed filters.

This method assures accurate register at the transmitter, but at the expense of a camera of restricted utility. If separate cameras are used for each primary color, great care must be exercised to secure precisely congruent scanning patterns in the three tubes. Identical scanning systems, fed in parallel if voltage-actuated or in series if current-actuated, must be used, and disturbing fields must be rigidly excluded. A similar register problem exists at the receiver.

Thus far, the only practical method of producing the received image is by projection of images from three picture tubes onto a common viewing screen. The picture tubes may employ white phosphors and be fitted with fixed filters having the trichromatic coefficients previously specified. Higher luminous efficiency may be achieved, however, by the use of colored phosphors in conjunction with additional "correcting" filters of high transparency. One difficulty with the use of colored phosphors is the tendency of the color to change with brightness.

Simultaneous color pictures of live subjects have been demonstrated using colored phosphors on 5-in. tubes projected through Schmidt optical systems onto a 9- by 12-ft screen.

79. Standards for the Simultaneous System. There is an evident advantage in choosing the same scanning standards for each channel in the simultaneous color system as are used for black-and-white transmissions, namely, 525 lines, 30 frames per sec,

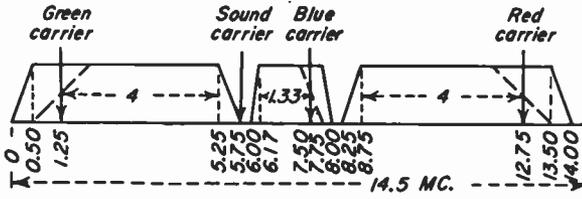


FIG. 72. Channel arrangement for a 525-line 30-frame simultaneous color system.

60 fields per sec, 4/3 aspect ratio. If this is done, the color system becomes compatible with the black-and-white system, and receivers designed for monochrome reception may receive the color transmissions in monochrome by tuning (with an r-f converter) to one of the simultaneous color transmissions. Since the green primary image contains nearly all the half-tone information of the image, reception of the green signal will produce a black-and-white image virtually indistinguishable from that produced by a black-and-white system televising the same subject. During the transition period from monochrome to color television, such a compatible system might prove of great value, since it would permit color transmissions to be received on receivers designed for monochrome service.

Although no definite proposals for standards for a simultaneous system have been made before the FCC, all present work on this system is predicated on compatible scanning standards, modulation methods, and channel spacings. A typical channel spacing for the simultaneous system is shown in Fig. 72. It will be noted that the green channel is identical to the standard channel for black-and-white transmissions. A noteworthy feature of this channel arrangement is the small band width occupied by the blue channel, the maximum v.f. being 1.33 Mc, or one-third that for the green-and-red channels. This reduction in the blue band width is possible because the visual acuity of the human eye is substantially less for blue light than for red and green lights. If the resolution of the blue image is reduced by a factor of 3 at normal viewing distances, no visible effect on the detail of the image is noted.

As a consequence of the narrow blue channel, the over-all band width for a simultaneous system, compatible with the black-and-white system, is 14.5 Mc. This compares with 16 to 18 Mc for a sequential system of the same resolution, flicker, and brightness performance.

A further reduction in band width in the simultaneous system, gained at the expense of some apparatus complexity, is the "mixed-highs" system shown in Fig. 73. In this system, the fine detail of the image is removed from the blue and red channels and combined with the fine detail of the green channel. The green channel then carries the full picture information and is available for compatible reception on black-and-white receivers. For color reception, the fine detail carried by the green channel is applied simultaneously to all three picture tubes, as shown. The fine detail is thus reproduced in shades of gray, whereas the larger details, corresponding to video fre-

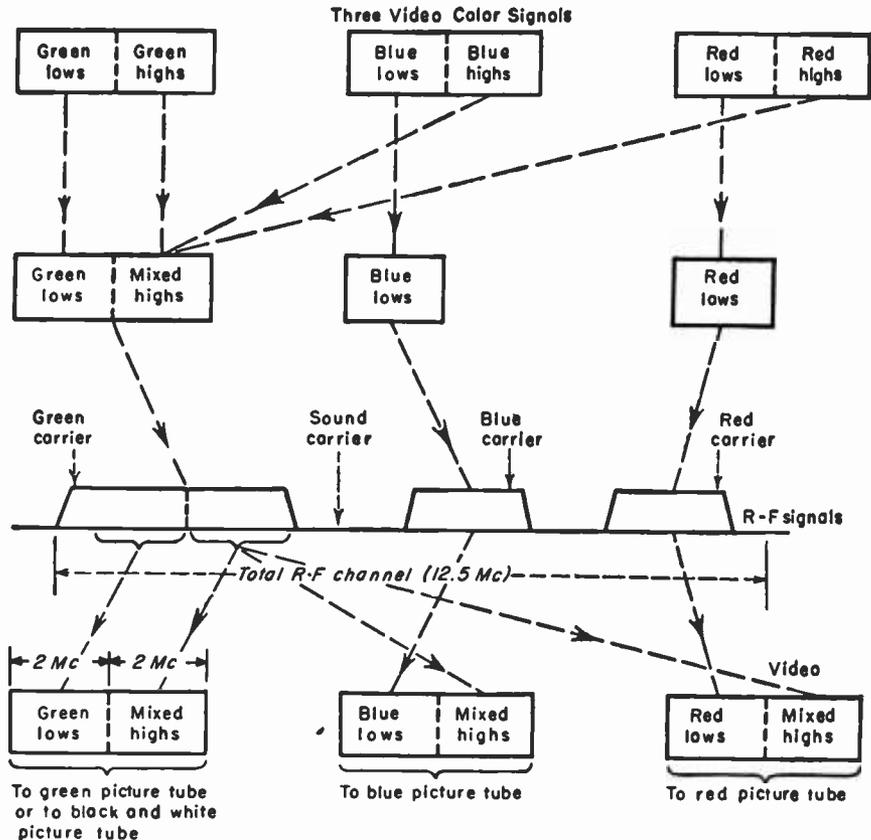


FIG. 73. Mixed-highs simultaneous color system.

quencies up to 2 Mc, are reproduced in color. The eye does not perceive the lack of color in the fine details, and substantially no difference is noticed, at normal viewing distances, when the mixed-highs connection is substituted for full simultaneous reproduction (full band width in each primary channel). The mixed-highs system permits a reduction in over-all band width to 12.5 Mc.

References

Color Television:

- Cherry, W. H.: Colorimetry in Television, *RCA Rev.*, 8, 427, September, 1947.
 Fink, D. G.: Color Television on Ultra-high Frequencies, *Electronics*, 19, 109, April, 1946.

- : Two Systems of Color Television, *Electronics*, **20**, 72, January, 1947.
 ———: Where Color Television Stands, *Electronics*, **19**, 104, May, 1946.
 Goldmark, P. C., et al.: Color Television, *Proc. IRE*, **30**, 162, April, 1942; **31**, 465, September, 1943.
 Kell, Fredendall, Schroeder, and Webb: An Experimental Color Television System, *RCA Rev.*, **7**, 141, June, 1946.
 Kell, R. D., et al.: An Experimental Simultaneous Color-television System (in three parts), *Proc. IRE*, **861**, August, 1947.
 RCA Laboratories: Simultaneous All-electronic Color Television, *RCA Rev.*, **7**, 459, December, 1946.

General Bibliography

Books:

- Fink, D. G.: "Principles of Television Engineering," McGraw-Hill, 1940.
 ———: "Television Standards and Practice," McGraw-Hill, 1943.
 Lohr, L. R.: "Television Broadcasting," McGraw-Hill, 1940.
 Maloff, I. G., and D. W. Epstein: "Electron Optics in Television," McGraw-Hill, 1938.
 Wilson, J. C.: "Television Engineering," Pitman, 1937.
 Zworykin and Morton: "Television," Wiley, 1940.

Periodicals:

- Beal, R. R.: Equipment Used in the Current RCA Television Field Tests, *RCA Rev.*, **1** (3), 36, January, 1937.
 Beers, Engstrom and Maloff: Some Television Problems from the Motion Picture Viewpoint, *J. Soc. Motion Picture Engrs.*, **32**, 18, January, 1939.
 Bingley, F. J.: VHF Multiple-relay Television Network, *Electronics*, **18**, 102, October, 1945.
 Bloecker, W. E.: Interconnecting Facilities for Television Broadcasting, *Electronics*, **20**, 102, November, 1947.
 Boothroyd, W.: U-h-f Television Relay System, *Electronics*, **20**, 86, March, 1947.
 Bradley and Traub: A New Television Projection System, *Electronics*, **20**, 84, September, 1947.
 Buzalski, T. J.: A Method of Measuring the Degree of Modulation of a Television Signal, *RCA Rev.*, **7**, 265, June, 1946.
 Cowan, F. A.: Coaxial Cable Networks, *Proc. IRE*, **35**, 1364, November, 1947.
 Deerhake, F. M.: 2000-Mc Television Program Chain, *Electronics*, **21**, 94, February, 1948.
 DuMont and Goldsmith: Television Broadcast Coverage, *Proc. IRE*, **32**, 192, April, 1944.
 Eddy, W. C.: Television Lighting, *J. Soc. Motion Picture Engrs.*, **33**, 41, July, 1939.
 ———: Television Studio Considerations, *Communication and Broadcast Eng.*, **4** (4), 12, April; **5**, 14, May; **6**, 20, June; **7**, 17, July, 1937.
 Engstrom, Kell, Bedford, Trainer, Holmes, Carlson, and Tolson: An Experimental Television System, *Proc. IRE*, **22**, 1241, 1246, 1266, November, 1934.
 Epstein and Pensak: Improve Cathode-ray Tubes with Metal-backed Luminescent Screens, *RCA Rev.*, **7**, 5, March, 1946.
 Fink, D. G.: Brightness Distortion in Television, *Proc. IRE*, **29**, 310, June, 1941.
 ———: Design Trends in Television Transmitters, *Electronics*, **21**, 76, January, 1948.
 ———: A Laboratory Television Receiver (in six parts), *Electronics*, Part I, *ibid.*, **11** (7), 16 July, 1938; Part II, *ibid.*, **11** (8), 26, August, 1938; Part III, *ibid.*, **11** (9), 22, September, 1938; Part IV, *ibid.*, **11** (10), 16, October, 1938; Part V, *ibid.*, **11** (11), 26, November, 1938; Part VI, *ibid.*, **11** (12), 16, December, 1938.
 ———: Photographic Analysis of Television Images, *Electronics*, **14**, 24, August, 1941.
 ———: Television Antennas for Apartments, *Electronics*, **20**, 96, May, 1947.
 ———: Television Broadcasting Practice in America, 1927 to 1944, *J. IEE (London)*, **92**, III, 145, September, 1945.
 ———: A Television Receiver for the Home, *Electronics*, **12** (9), 16, September, 1939.
 Friend, A. W.: Television Deflection Circuits, *RCA Rev.*, **8**, 86, March, 1947.
 Gannett and Green: Wire Transmission System for Television, *Bell System Tech. J.*, **6**, 616, October, 1927.
 Goldmark, P. C.: A Continuous Type Television Film Scanner, *J. Soc. Motion Picture Engrs.*, **33**, 18 July, 1939.
 ———: Problems of Television Transmission, *J. Applied Phys.*, **10**, 447, July, 1939.
 ——— and Dyer: Quality in Television Pictures, *Proc. IRE*, **28**, 343, August, 1940.
 Gray, Horton, and Mathes: The Production and Utilization of Television Signals, *Bell. System Tech. J.*, **6**, 560, October, 1927.
 Hanson, O. B.: Experimental Studio Facilities for Television, *RCA Rev.*, **1** (4), 3, April, 1937.
 Ives, H. E.: Television, *Bell System Tech. J.*, **6**, 551, October, 1927.
 ———: Transmission of Motion Pictures over a Coaxial Cable, *J. Soc. Motion Picture Engrs.*, **31**, 256, September, 1938.
 Kallman, H. E.: Gradation of Television Pictures, *Proc. IRE*, **28**, 170, April, 1940.
 Maloff, I. G.: Electron Bombardment in Television Tubes, *Electronics*, **17**, 108, January, 1944.
 ——— and D. W. Epstein: Reflective Optics in Projection Television, *Electronics*, **17**, 98, December, 1944.
 Markus, J.: New York-Boston Microwave Television Relay, *Electronics*, **21**, 114, January, 1948.
 Mautner, R. S., and O. H. Schade: Television High Voltage Power Supplies, *RCA Rev.*, **8**, 43, March, 1947.

- Morris and Shelby: Television Studio Design, *RCA Rev.*, **2** (1), 14, July, 1937.
- RMA Staff: Television Resolution Chart, *Electronics*, **20**, 123, December, 1947.
- Roe, J. H.: New Television Field Pick-up Equipment Employing the Image Orthicon, *Proc. IRE*, **35**, 1532, December, 1947.
- Schade, O. H.: Magnetic Deflection Circuits for Cathode-ray Tubes, *RCA Rev.*, **8**, 506, September, 1947.
- : Radio-frequency-operated High-voltage Supplies for Cathode-ray Tubes, *Proc. IRE*, **31**, 158, April, 1943.
- Schlesinger, K.: Magnetic Deflection of Kinescopes, *Proc. IRE*, **35**, 813, August, 1947.
- Seeley and Kimball: Transmission Lines as Coupling Elements in Television Equipment, *RCA Rev.*, **3** (4), 418, April, 1939.
- Striely, M. E.: Coaxial Cable System for Television Transmission, *Bell System Tech. J.*, **17**, 438, July, 1938.
- and Weis: Television Transmission, *Proc. IRE*, **29**, 371, July, 1941.
- Television Transmitters. *Electronics*, **12** (3), 26 March, 1939.
- Wendt and Fredendall: Automatic Frequency and Phase Control of Synchronization in Television Receivers, *Proc. IRE*, **31**, 7, January, 1943.
- Wright, A.: Television Receivers, *RCA Rev.*, **8**, 5, March, 1947.
- Zworykin, V. K.: The Iconoscope, a Modern Version of the Electric Eye, *Proc. IRE*, **22**, 16, January, 1934.
- : Television, *J. Franklin Inst.*, **217**, 1, January, 1934.

CHAPTER 20

FACSIMILE

By R. E. MATHES,¹ B.S.

1. General Considerations. The term *facsimile* has been applied to that branch of the science of graphic electrical communication which endeavors to convey the physical form, and even the light shadings of the original subject matter. Such information cannot be instantaneously or simultaneously transmitted, and it is thus necessary to do so bit by bit sequentially. The manner of doing this effectively is to divide the original into a large number of elemental areas and to transmit signals to indicate the relative light shades of these areas. Such shades are then reproduced more or less accurately at the receiver. The elemental areas are recorded in the same sequence, thus building up the record similar to the building of a brick wall.

The accuracy of reproduction depends upon the number of these elemental areas in the picture. It makes no difference as to the size of the finished record; the resolving power is entirely a matter of the number of elemental areas satisfactorily transmitted. It takes just as many tiny areas to represent well a face on a postage stamp as it does to represent a well face in larger areas on a 10-ft enlargement.

To transmit sequentially and to record these areas with necessary fidelity requires highly accurate mechanisms, synchronizing means, communication circuits, amplifier circuits, and scanning and recording devices. Such means have been the subject of intensive development efforts for many years.

2. Historical. In 1842 Alexander Bain proposed a facsimile system which had in it all the pertinent functions included in the most modern derivations. Figure 1, a copy of his original system, shows the transmitting arrangement clearly. Synchronous action was afforded by the pendulums at transmitter and receiver, line advance by moving the message plates upward a short distance at the end of each swing of the pendulum, and elemental area scanning by the contact of the metal brushes. Caselli produced an improved system in 1865, Korn another in 1902, Belin in 1920. The American Telephone and Telegraph Company opened a public service in the United States in 1925, and the RCA inaugurated public service with London in 1926. An excellent bibliography covering this growth is given by J. L. Callahan.²

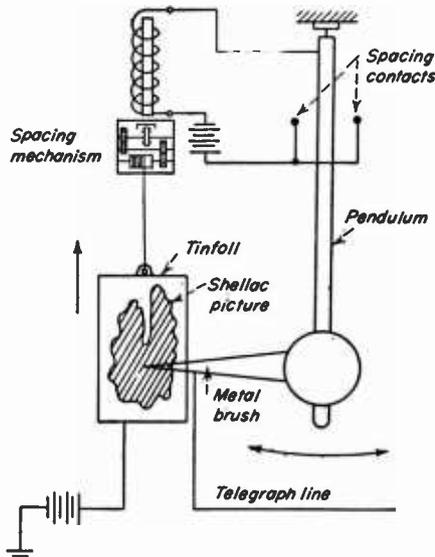


FIG. 1. Alexander Bain's original picture apparatus.

¹ Finch Telecommunications, Inc.

² Callahan, J. L., A Narrative Bibliography of Radio Facsimile, "Radio Facsimile," RCA Institutes Technical Press, 1938.

3. Transmission. The functions necessary to transmit a facsimile record are as follows:

1. A scanning system to explore the elemental areas of the subject and identify their individual light shadings in terms of an electrical current.
2. A modulation circuit to provide this fluctuating current in a form suitable for transmission over the communication system available.
3. A mechanism to provide an orderly exploration of these areas by the scanning system.
4. An electrical drive system to run the mechanism at a uniform, predetermined rate, within close tolerances.

4. Scanning. Modern methods invariably use an elemental area of intense light, either transmitted through or reflected from the original subject and picked up by a phototube. The diaphragm is sometimes placed at point X-X after the pickup lens rather than at the condenser lens. The effective light of the scanning system is proportional to the product of the intrinsic intensity of the light source and the solid angle subtended by the objective lens at the surface of the subject.

Ideally, the elemental area will be of infinitesimal width. This cannot be realized practically, and therefore all scanners have an effective light spot of finite width. This gives rise to a distortion

known as the *aperture distortion* which modifies the electrical signals so they are not a true representation of the instantaneous changes of the shadings or tonal values of the subject.

In Fig. 3 a light spot of a width nearly as great as that of the finest vertical bars is shown at *a* and the resultant electrical response at *b*. Such relatively great finite width of the spot produces a trapezoidal wave form which becomes triangular as the spot width becomes just equal to that of the vertical lines of the subject. On the other hand, a narrower spot, such as *c*, will produce a wave form *d*, which, although still trapezoidal, approaches a true rectangular or "square" wave shape as the spot width approaches zero.

The narrow spot will permit the interpretation of more detail of the subject but will result in an electrical wave form which has pertinent and necessary component frequencies (harmonics) considerably higher than those produced by the wider spot. In other words, the aperture distortion has an action approximately equivalent to that of a low-pass filter having a perfect phase characteristic. Figure 4 shows the manner in which the amplitude of the harmonic components decreases for three different apertures.¹ The curve for the rectangle corresponds to an aperture of infinitesimal width; that for the trapezoid corresponds to an aperture whose width is two-thirds that of the vertical line to be scanned (Fig. 3); that for the triangle corresponds to an aperture whose width is equal

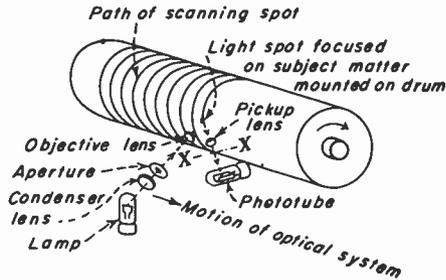


FIG. 2. Necessary elements for scanning.

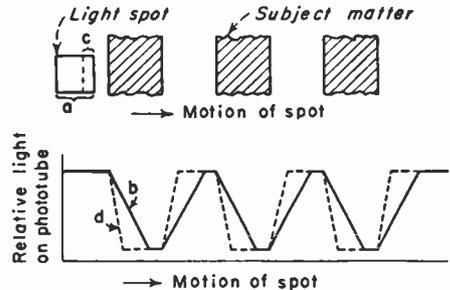


FIG. 3. Distorting effect of finite width of the scanning light spot or aperture. *a* and *b* spot width = three-fourths width of narrowest line to be scanned. *c* and *d* spot width = one-fourth width of narrowest line to be scanned.

¹ Jolley, L. B. W., "Alternating Current Rectification," Wiley, 1928.

to that of the vertical line. It will be seen that both the triangle and trapezoid components drop to a negligible value very quickly, whereas the value of rectangular components does not drop to 2 per cent until about the twenty-ninth harmonic. The peak amplitude of these shapes were all taken as 1.0.

The practical determination of the detail required depends upon the use to be made of the record. However, commercial experience to date teaches that, for an ultimate enlargement of the recording by not more than 4 to 1 over the original, a texture of 100 to 120 scanning lines per inch is ample. One system uses a texture of 200 lines per inch and this, of course, permits of still greater enlargement of the recording. If detail is set as equal in both dimensions, a minimum width of vertical lines is indicated as about 0.008 in. It is therefore necessary merely to produce a sufficiently accurate representation of a line of this width on the record. Practically, considerable inaccuracy may be permitted because the minimum detail the average eye can differentiate is that area which subtends an angle at the eye of 0.00065 to 0.00070 radian (approximately 2 min). It is found that, for photographic recording, a rectangular aperture 0.006 in. wide is a practical compromise. For so-called "visual"

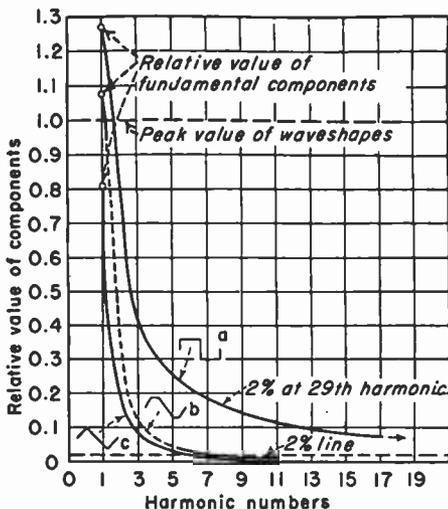


FIG. 4. Relative value of harmonic content of signal produced by various width apertures: a for square wave shape from aperture of infinitesimal width; b for trapezoidal wave shape from aperture two-thirds width of narrowest line; c for triangular wave shape from aperture width equal to narrowest line. Also note relative value of fundamental component for each.

methods of recording which have limited definition capabilities in themselves, a round aperture with diameter equal to the width of the vertical lines is sufficient. Such compromise choices of practical aperture size will reduce the pertinent frequency components as much as is permissible.

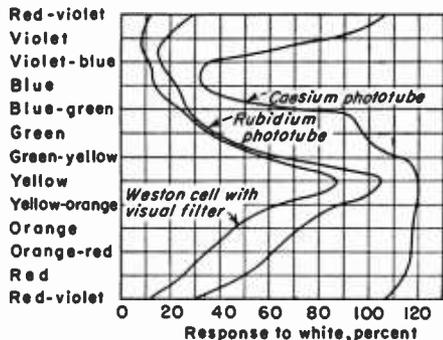


FIG. 5. Sensitivity of several phototubes as a function of color of incident light.

Another phase of scanning is the spectral distribution of the light reflected from the surface of the subject, as referred to the distribution of the phototube sensitivity. Figure 5 shows curves of such distributions. The curve for the rubidium tube fairly closely follows that of the eye; i.e., it is about equally panchromatic. Use of this type tube will give a black-and-white recording in which the various colors, as well as the tonal values, are given a weighting closely approximating that assigned by the eye and results in a more effective

reproduction, even though the original was in color. This is of importance for facsimile service because the system should be capable of handling any type of subject matter that may be submitted. Unfortunately, the sensitivity of rubidium is low, and

this type tube can be used only where there is an excess of light available to the phototube. Therefore the cesium-surfaced phototube is the type more generally used.

For the basic cesium-type phototube there have recently been developed several photosensitive surfaces having different spectral responses. Practice is not uniform as to selection of the most suitable response. Some designers choose cells with maximum response toward the ultraviolet, others choose the maximum toward the infrared, while still others strive for a uniform response over the visible spectrum. Some of the latter use photographic-type glass or gelatin filters to modify the cell response in an effort to achieve true color blindness. The latter practice is wasteful of the available light and should be used only when color blindness is essential, *i.e.*, when colored subject matter as well as black-and-white subjects must be scanned, but recorded as black and white.

The scanning system usually includes the phototube and its immediately associated amplifiers. The light intensity available at the phototube is very low—on the order of 0.001 to 0.005 lumen—and the voltage output of this tube is likewise low. It can be increased by increasing the load resistance, and this has sometimes been made as high as 10 to 25 megohms. However, the interelectrode capacitance of the tube becomes serious when slanted across such high value, even though these capacitances are of the order of 1 to $2.5\mu\text{f}$. The effect is that of a low-pass filter to limit the higher frequency components and thus limit the possible scanning speed of the facsimile

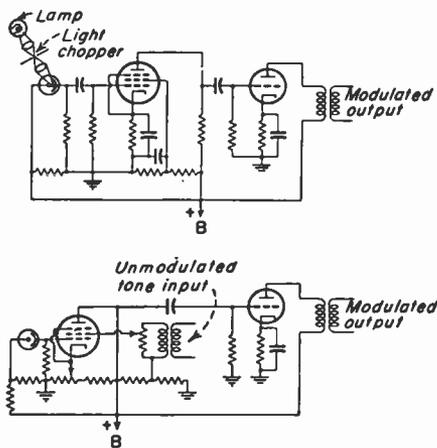
equipment. A gas-filled phototube operated at normal voltages is several times more sensitive than a vacuum phototube, but also gives the effect of a low-pass filter and thus discriminates against the fine picture details. This is caused by the finite time required for the gas tube to deionize.

Use of lower phototube load resistance— $\frac{1}{2}$ to 1 megohm—to permit of sufficiently h-f response for higher speed operation makes it difficult to amplify the weak output. The simplest way would be to use direct resistance-coupled amplifiers. However, the variations in voltages, emission, and contact potential are of the same order of magnitude as the desired signal and in such amplifiers are superposed on the signal. One solution for this difficulty is to modulate the light beam at an a.f. and use conventional a-c amplification. The modulation can be applied directly to the lamp in the case of a gaseous light source, such as neon, helium, or mercury vapor, or it may be accomplished by cutting the beam by a mechanical chopper, such as a string galvanometer, vibrating reed, or chopper disk.

FIG. 6. Typical scanner amplifier circuits. (Above) A-c amplifier for use with light chopper in the optical system. (Below) Modulator amplifier for use with d-c output of phototube.

Another solution is to apply the phototube output to a sensitive modulator, such as a balanced bridge circuit, and to amplify the resulting modulated tone. Figure 6 (above) shows a typical arrangement of the first type and Fig. 6 (below) that of the second type.

In either of these types it is essential that the audio tone which acts as the carrier for the facsimile modulation be of a high enough frequency that the shortest signal to be sent (*e.g.*, a line 0.008 in. wide) be composed of enough tone cycles to form a sufficiently accurate envelope.



5. Modulation. The signals produced by the scanner may be transmitted over different types of communications systems, and there are numerous ways in which the signals may be applied as modulation for these systems. Those of present commercial importance will be outlined briefly:

1. For radio circuits:
 - a. Subcarrier a.m.
 - b. Subcarrier f.m.
2. For landlines:
 - a. Double-side-band subcarrier a.m.
 - b. Single-side-band subcarrier a.m.
 - c. Subcarrier a.m.
3. For oceanic cables:
 - a. D-c transmission

For radio circuits, subcarrier a.m. is presently standardized for facsimile broadcasting in the 88- to 108-Mc band. It may also be utilized on v-h-f and u-h-f radio circuits which are operated with f.m. or pulse modulation, or are otherwise rigidly limited or

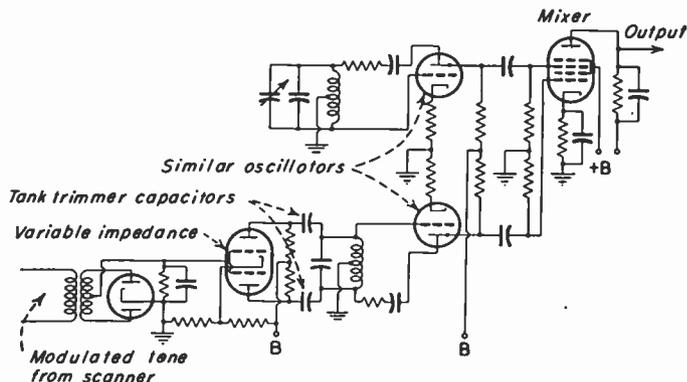


FIG. 7. Diagram of transmitting converter for subcarrier f-m method.

controlled as to amplitude variations. It is the most straightforward modulation method.

For radio circuits subject to level variations, such as long distance h-f or short-distance m-f circuits, facsimile signals should be by subcarrier f-m modulation to provide for the required limiting. The modulation process is essentially this: The input from the scanner is rectified and applied to a push-pull triode which, in turn, acts as a variable resistance in series with balanced trimmer capacitors connected across the tank circuit of an i-f oscillator. The variation in resistance of the triode varies the effectiveness of the trimmer capacitor on the natural period of the tank circuit, and thus provides f.m. in accordance with the light variations of the scanned subject matter. A second i-f oscillator of fixed frequency is beat against the first and the difference in frequency taken as the output as in the conventional beat-frequency oscillator. The result is an audio output which is frequency-modulated over a relatively large percentage of the audio mid-frequency. This output is applied to a radio telephonic transmitter as a.m.

The principle of f.m. can, of course, be applied directly to the r-f carrier instead of through the medium of a subcarrier, but this requires specialized equipment at both the radio transmitter and the radio receiver, which is not required in the above scheme. Furthermore, it does not overcome any variations in the audio equipment or on the control lines as does the subcarrier method. Phase modulation can also be used if the radio propagation conditions permit.

For *landlines* the standard procedure is to transmit the a-m subcarrier tone, *i.e.*, the signal output of the scanner. Necessary amplifiers are used to provide the desired level and impedance matches to the telephone lines which are used as the transmission circuits. Most systems now couple to the lines directly through standard repeat coils.

Both double- and single-side-band transmission is extensively used. In both cases care is taken not to utilize the frequencies below about 1,000 cycles because of the inherent poor phase characteristic of wire lines at the low frequencies. In the case of single side band it is usually the upper side band that is suppressed. The frequency of the carrier tone used varies in different systems from about 1,800 to 5,000 cycles. The exact value chosen is usually dependent upon the h-f characteristics of the line or channel to be used. It is made as high as is possibly consistent therewith.

For *Oceanic Cable*. A new method for the transmission of pictures on the trans-oceanic cables was put into service in 1939. This system is the only practical system which transmits the picture signals directly without any intermediary modulation or carrier. The phototube current is built up by d-c amplifiers and applied to the cable in that form.

Recently some use has been made of subcarrier f.m. on telephone toll circuits to minimize the effect of level variations on this type of line which is not so carefully controlled as special more expensive lines provided particularly for facsimile (or wire-photo) service.

6. Mechanisms. In order that the elemental areas be scanned sequentially, it is necessary that mechanical means be provided to move the scanning light spot relative to the subject matter in such a manner that the entire area to be transmitted is covered in a predetermined order. Many different sequences have been proposed in the past, *e.g.*, scanning alternate lines, diagonal and crisscross patterns, etc. However, the simple uniform scanning across the width of the subject, with line-by-line advance along the length of the subject, is the easiest and the most readily adaptable to various mechanisms. The relative motion may be obtained by moving either the light spot or

the subject, or both. The latter is the more usual, in which the light spot is moved along one dimension of the subject and the subject itself is moved along its other dimension.

Figure 8 shows the most popular scheme, in which the subject is wrapped around a drum which revolves relatively rapidly in front of an optical and phototube pickup assembly. This assembly is carried on a track and caused by a lead screw to move along parallel to the axis of the

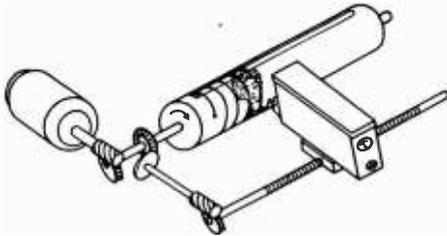


Fig. 8. Drum type of machine.

drum at a relatively slow rate. This rate is so chosen by proper gear ratios that it will travel axially exactly the width of one scanning line for each revolution of the drum. In some designs the optical and pickup system is mounted in a fixed position, and the drum is slowly advanced along its axis as well as being revolved rapidly.

Figure 9 indicates a different arrangement in which the optical system is comprised of two, three, or four identical lens systems for projecting the light spot on to the subject matter and for picking up the reflected light. These are mounted concentrically like spokes of a wheel and are revolved rapidly. The subject matter is applied face down to a one-half, one-third, or one-fourth section of cylindrical tubing which is mounted concentric with the optical system. The subject is moved slowly along the cylinder by a belt or other device. This scheme has the great advantage that it can handle any size or thickness of subject matter so long as but one dimension does not exceed the peripheral length of the cylindrical section. Also successive subjects can be fed to the machine without need for stopping the machine or, alternatively, providing removable drums and rather complicated clutch mechanisms to prevent loss of

synchronism. However, it has the severe handicap—yet to be fully overcome—of requiring an ultrafine degree of mechanical precision because all the optical systems must be exactly equal in optical efficiency as well as track perfectly.

The precision of the mechanisms must be of high order in any good facsimile system, but the exact requirement is practically dependent upon the definition of which the recording system is capable. The photographic method is the best from this standpoint, and in this case an instantaneous hunting, as between successive lines, of more than 0.08 deg of drum rotation is a serious error. Likewise, inaccuracies in the lead screw, such as spacing variations between successive lines of 0.001 in. are quite noticeable in the record—particularly if the variations are of a periodic nature occurring every 2 to 10 or so lines. drum or motor, minute eccentricities, voltage fluctuations, all must be carefully

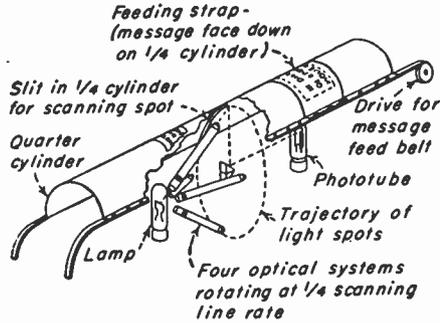


FIG. 9. Quarter-cylinder type of continuous feed machine.

Gear ripples, dynamic unbalance of the gear, voltage fluctuations, all must be carefully guarded against.

Often cam and contact devices are applied and lined up with the interval on the drum between the trailing and leading edges of the subject, or "phasing line." These are used for special functions such as sending a synchronizing or level-control signal. They are usually carried on either the drum or driving shaft.

Many commercial or military applications require a portable transmitter. Particularly is this true for news picture

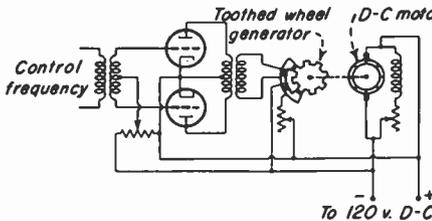


FIG. 10. Diagram of "Hammond brake" speed control for d-c motors.

work where the source of the subject matter may be anywhere at any time. Several designs of portable equipment have been produced for mounting small, comparatively light units in carrying cases. However, it is most difficult to get and retain the same precision as in equipment permanently installed.

7. Electrical-drive Systems. The mechanism may be variously driven by controlled d-c motors, 60-cycle synchronous motors, low-speed phonic wheels, and, lately, by the Alnico type of toothed wheel which operates directly at some 75 rpm from 60 cycles, 120 volts a.c., or they may be designed to operate at a speed such as 100 rpm from a drive frequency of 160 cycles. These may be driven from ordinary vacuum-tube power amplifiers. Some use one large motor to drive both the transmitting mechanism and a local monitor recorder, and others use two small motors, one for the drum and the other for the lead screw. One scheme of

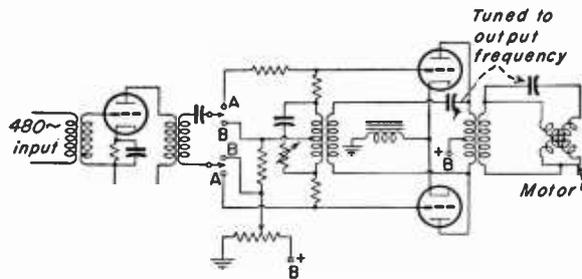


FIG. 11. Controlled thyratron inverter for driving synchronous motors.

directly at some 75 rpm from 60 cycles, 120 volts a.c., or they may be designed to operate at a speed such as 100 rpm from a drive frequency of 160 cycles. These may be driven from ordinary vacuum-tube power amplifiers. Some use one large motor to drive both the transmitting mechanism and a local monitor recorder, and others use two small motors, one for the drum and the other for the lead screw. One scheme of

controlling a d-c motor which has been used for years is the Hammond brake scheme, shown in Fig. 10, and a scheme for driving a standard fractional-horsepower synchronous motor is a push-pull thyatron inverter circuit, such as shown in Fig. 11.

RECEPTION

For purposes of reception there must be the following:

1. A recording system which will translate the signals into visual markings
2. A sensitive surface to receive such markings
3. A mechanism to provide an orderly relative motion of the record surface
4. Necessary filters, amplifiers, rectifiers, etc.
5. Synchronizing means to maintain the mechanism in instantaneous phase with the transmitter

8. Photographic Recorders. The system most used records photographically on film. It has the advantage of giving the best detail and quality of definition and also readily provides a negative from which further processing, such as newspaper reproduction, can best be accomplished. It is usual to employ a thin base film that can be easily handled on a drum. The sensitive emulsion should be chosen for linearity of light response rather than extreme contrast. It should develop rapidly, and special films have been made which can be developed, fixed, and dried quickly so the picture can be used as soon after reception as possible.

Recorders take many forms. Glow tubes of neon, argon, helium, or a mixture of gases are formed so the glow takes place in a crater, thus confining the glow and giving essentially a point source of light of fair intrinsic brilliance. In this tube the light intensity follows directly the modulation of the signal. Very recently a "concentrated-arc" lamp utilizing a zirconium cathode has been announced. It is also controllable and can be modulated and has a very high intrinsic intensity.

Other schemes use a local light source of fixed amplitude, such as filament or arc lamps of various types. The light is then passed through a modulating device before it strikes the film. One method sometimes used in Europe is the Kerr cell coupled with Nicol prisms to polarize the light.

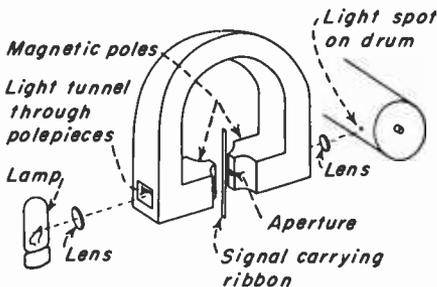


FIG. 12. Light-valve optical system.

If the ribbon is mounted parallel to the axis of the drum, this results in "variable-density" recording. If mounted perpendicular to the axis, it results in "variable-width" recording. Both methods are similar to those used in recording sound on film in the motion-picture field. For facsimile the former is by far the better.

A fourth scheme utilizes a D'Arsonval galvanometer or some modification. In this case a small mirror is mounted on the signal-carrying wires or coil set in the magnetic field. The signal causes an angular movement of the mirror and so changes the orientation of the reflected light on a fixed aperture. This is diagrammed in Fig. 13, which also shows a linear penumbra which permits variable-density recording without critical adjustments of the galvanometer or signal level.

9. "Direct" Recording. Various schemes have been worked out for methods of "direct" recording in which the record appears almost instantaneously and need not be developed and fixed as with a photographic film.

Hot air has been used to discolor a presensitized paper, or to evaporate an opaque coating to permit the color of the paper base to show through. The air stream was keyed or triggered by the signal. An alcohol ink has been vaporized and blown on a glossy paper on the drum. It was keyed by a shutter on a signal relay.

One group of commercial equipment, which is growing in its usage by the telegraph industry, uses a wire stylus of small contact area, bearing on a white-coated dry paper. The signal current passes through the paper from the stylus to the drum, burns off the white coating, and permits the black paper base to show through. The scheme is shown in Fig. 14.

Another important field of direct recording is that of electrochemical action. Certain groups of molybdenum and other salts will change color when current passes, as will also organic azo- and pyrocatacin dyes. The process is sensitive and very rapid, but difficulty is experienced in obtaining a stable reaction that will hold for a matter of years. The recording proper is stable, but the unreacted

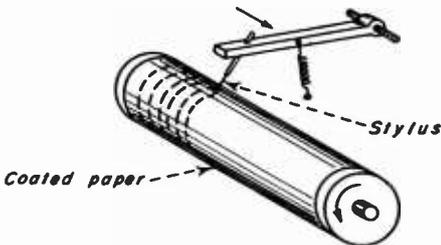


FIG. 14. Elements of stylus recording on coated, carbonaceous paper.

Much work has also been done to utilize carbon recording. In this process ordinary cheap carbon paper is laid with its face against plain white paper and the combination advanced slowly between a rapidly rotating drum carrying a raised spiral of wire and an axially parallel bar. The bar is driven by loud-speaker-type magnets to vary the pressure between it and the spiral, in accordance with the signal. This method is best adapted to black-and-white recording but can be used to record linearly half-tone pictures if the amplifiers are carefully compensated for the nonlinear pressure-density characteristic of the carbon paper.

10. Recording Sheets. Some slight use is made of photographic recording of a positive directly on paper, but the majority of the work is negative recording on film stock which is itself delivered, or from which a delivery print is made. The emulsion is chosen to match as nearly as possible the spectral distribution of the recording lamp. Within this requirement it is also desirable that the emulsion have a good linear region, that the gamma be fairly high to require a lesser intensity range of the recording spot, and that it be relatively color-blind, if possible, to permit working in the darkroom under safelights. For black-and-white recording, a highly contrasty and sensitive film is desired and a commercial process film is used. For linear recording a less contrasty and more sensitive film is best, and special emulsions have been made available which lie approximately between an orthochromatic and a panchromatic as regards color response.

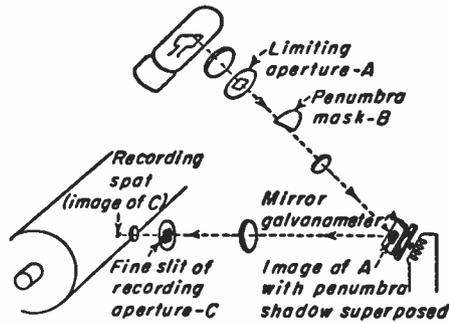


FIG. 13. System with linear penumbra mask and mirror galvanometer.

chemicals cause the background to discolor with age, particularly if exposed to light. Satisfactory recordings have been made at speeds in excess of 1,500 fpm, linear spot travel. The inherent definition ability of this method at present is much less than with photographic recording but is ample for certain services, such as message handling or home recording. Present proposals for home reception of facsimile broadcasts on f.m. all contemplate its use. Figure 15 shows the arrangement.

11. Recording Mechanisms. The recording system utilizes mechanisms of the same general type as for transmission, some of which have already been indicated, *e.g.*, in Fig. 8. In this case, however, the mechanisms must be held to even closer tolerances than in the transmitters in order that deviations be not apparent in the record. Because of its greater inherent definition the photographic process requires the most precise mechanisms. Specific tolerances are indicated in Table 1.

The base paper used for opaque coatings is fairly heavy, carbonaceous, coarse-grained. It is usually coated with titanium oxide to give a light gray or off-white shade before recording. The paper for electrochemical recording must be of high quality, chemically free of metallic impurities, and, in particular, must have high wet strength to withstand the feeding and scuffing stresses of the recorder while it is wet.

In particular, the lead screw must be very accurate in order that irregular feed

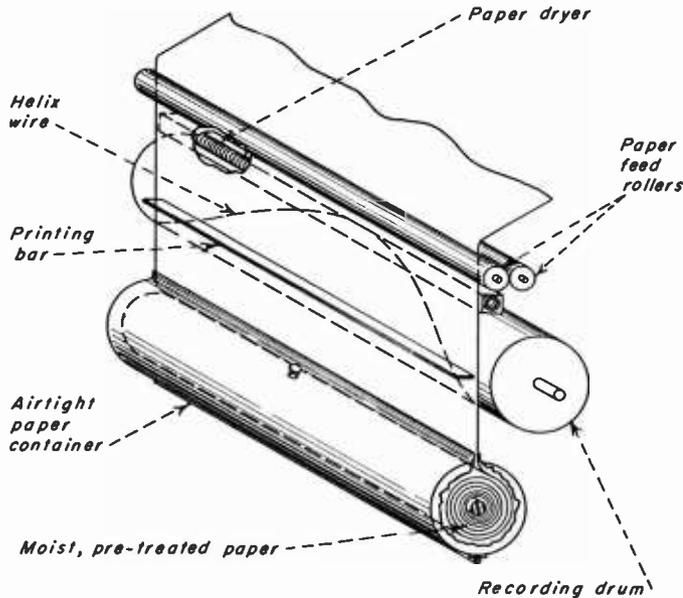


FIG. 15. Elements of spiral and bar recording on electrochemical paper.

between adjacent lines does not result in light or dark streaks in the recording. Also the drum or equivalent must travel smoothly and without hunt within the stroke or from stroke to stroke. This infers complete lack of gear ripple and a drive motor with a uniform instantaneous rate of rotation. These may usually be minimized by reducing the mass and inertia of the drum system or by use of flexible coupling and flywheel action to filter out these quick-speed variations.

Many receiving machines are fitted with cams for the purpose of using quasi- or full-automatic synchronizing or phasing. This is particularly necessary in home recorders or in a large network receiving a picture simultaneously in several offices. Most of these utilize the interval between the end of one scanning line and the beginning of the next as such a phasing signal. One variation is in the start-stop type of reciprocating machine, in which the travel of the stylus is arrested at the end of each complete stroke cycle and is then released for the next cycle by receipt of the phasing signal.

Many appliances have been developed, such as sheet folders and cutters, recordings on both sides of the record sheet, automatic drum loaders and discharge schemes, automatic hold circuits to permit reruns if needed, etc. Some, notably in the international

Table 1. Operating Standards of Typical Facsimile Systems Used on Wire Lines in the United States

	System A	System B	System C	System D	System E
Drum speed, rpm...	180	100	90	100	90
Scanning lines per in.....	100	100	96	100	96
Max picture size...	7¼ × 11¼	8¼ × 9	8 × 8	7 × 9	8 × 9½
Frequency of standard fork, cycles...	60	60	1,800	60	
Phasing method...	Automatic—from synchronizing dash sent at start of each picture				
Carrier frequency on line, cycles	2,150 on trunks 2,500 on local loops	1,920	1,800	2,400	1,800
Modulation transmitted	Double side band	Double side band	Double side band	Double side band	Double side band
Equalization.....	Amplitude equalizers along line and shaping discriminators at terminals	Permanent network equalized by phone company	No equalizer, use 1,000-cycle high-pass filter	No equalizer, use single K-section, 1,800-cycle, high-pass filter	Nothing
Max picture modulation, cycles....	1,250	950	650	1,000	400 (approx)
Permissible skew, in.....	½	⅓	¼	⅓	
Accuracy of synchronizing frequency	1 in 100,000	1 in 1,000,000 or better	1 in 500,000	1 in 100,000	
Required stability of line level, db	Uses pilot channel for AGC	0.1	3	¼
Type line system..	Own lines	Network leased	Long-distance toll calls		

field, are equipped with gear shifts to permit meeting various conditions and standards of other countries.

Machines have often been designed so they may readily be used as either transmitters or recorders. Others have been designed for monitor service in conjunction with transmitters.

12. Receiving Circuits. Most of the facsimile systems, either wire line or radio, record the signals in essentially the same general form as they are received. For these the received signal is amplified, rectified, and applied to the recorder either directly or through some type of vacuum-tube keyer. Invariably the frequency band is limited to that just necessary to pass the pertinent signal components. This is usually accomplished by the use of filter networks; the process is carried to a more or less high degree in the variously designed systems. Its purpose is primarily to increase the effective signal/noise ratio, but on long wire lines or extensive networks it also eliminates those lower frequency signal components whose phase would be sufficiently distorted to have a deleterious effect on the recorded copy.

Some few systems use special circuits at the recorder. In order that the recorded densities of the copy have a satisfactorily linear relation to the densities of the original subject, it is necessary that the amplitude characteristic of the amplifier be predistorted in a curve conjugate to that of the recorder characteristic. Of course, this can, and is, often done at the transmitting end instead of at the receiver. The present tentative RMA standards for facsimile broadcasting require a logarithmic response at the transmitter approximately to correct for the response curves of the various electrochemical recording papers in use.

A second system requiring special "signal shaping" is the method for transmission by submarine cable. This is more fully treated in Sec. 17.

A third such system is the subcarrier f-m method. The heart of this method lies in the special treatment accorded the signal at the recording location. In this case the

signal is amplified to a usable value, say zero level, and then applied to a very rigid amplitude limiter, *e.g.*, one showing no change in output level for a variation of the input on the order of 70 db. The output is then applied to a frequency-discriminating network, *e.g.*, cutoff slope of a low-pass filter. This reinserts an amplitude variation which is proportional only to the variation of frequency of the received audio subcarrier. Thereafter the signal is handled as an ordinary a.m. The system is shown in Fig. 16.

Several f-m receiver manufacturers use ratio detectors in lieu of r-f or i-f limiters. These receivers are satisfactory for reception of subcarrier a-m facsimile signals only in locations of high signal level or where r-f signal level does not vary. For other local-

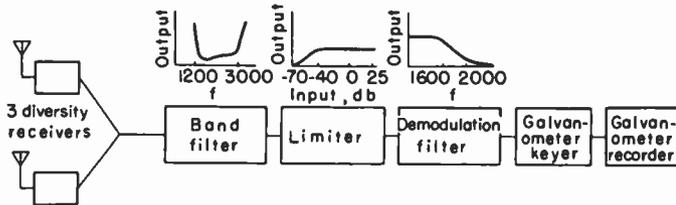


Fig. 16. Block diagram of receiving system for subcarrier f-m method.

ities, subcarrier f.m. is desirable to give a sufficient limiting action to minimize streaking of the recordings.

SYNCHRONIZING

In order that the recording may be properly built up to appear to be the equivalent of the original, it is necessary that the elemental areas be recorded in the same geometric relation with the others as are the elemental areas scanned at the transmitter. This demands that the transmitting and recording mechanisms operate in practically perfect synchronism and phase (or frame) with each other. It is now customary to drive the mechanisms by synchronous motors or by accurately speed-controlled d-c motors. The phasing is accomplished by manually or automatically causing the recorder to start a scanning line simultaneously with the start of a scanning line at the transmitter. This is done either at the start of a subject or periodically throughout the reception.

13. Drive Motors. For home recordings ordinary 60-cycle synchronous motors are often used. These are satisfactory if the transmitter and recorder are connected to the same power network or to networks that are interconnected or synchronized with each other. However, this is a considerable limitation, and the trend appears to be to transmit a control frequency to the recording station and to use that to control or drive the motor.

The schemes for driving or controlling the speed of the motors on the recorders are identical to those used on the transmitters, as mentioned in Sec. 7. All of them require a source of standard frequency for controlling the driving circuits. The great majority of present systems use independent tuning forks at the transmitter and recorders, or else they connect directly to a-c power lines which are interlocked as to frequency. However, this is a considerable limitation, and the tendency is toward the use of the phasing signal to control or correct a local oscillator which, in turn, drives or controls the motor.

14. Speed Controls. Many fixed installations for point-to-point communication—and most systems using photographic recording—obtain their speed control from tuning-fork oscillators. The frequency may lie between 60 and 1,500 cycles.

Some take elaborate precautions to enclose the forks in heated and heat-insulated chambers, with the temperature variations held to 0.01°C or less. The drive circuits are carefully engineered to minimize or eliminate variations in the drive due to supply voltage, tube emission, load, or similar variations. Recently it has been realized that

changes in atmospheric pressure have a considerable effect on the frequency stability of forks.

Other systems utilize a fork which is compensated for temperature and therefore does not require careful temperature control. Such forks are made of a bimetallic layer structure in which the two metals have opposite temperature coefficients. By proportioning the two metals properly, the resultant for the fork can be made stable to within 0.05 ppm per deg centigrade. However, its frequency is susceptible to variations in drive and atmospheric pressure.

15. Phasing Methods. In start-stop systems the recording gearing is chosen so the scanning-line cycle is traversed faster than at the transmitter, and the motion is arrested at the end of each cycle. The transmitter then sends a release or start signal at the commencement of each new scanning-line cycle, which releases the drive at the recorder. Thus the recorder is in step with the transmitter at the commencement of each line. Figure 17 shows one start-stop arrangement.

In other systems a phasing signal is sent at the start of each transmission schedule, and this cooperates with a cam on the recorder to slow down or speed up the drive until the two machines are accurately phased. This circuit is then disabled, and the relative equality of the fork frequencies at the two stations is depended upon to retain the phasing. In still others a special signal is sent which releases a clutch at the recorder and thus starts it in phase with the transmitter. This is essentially the same procedure as is used in the start-stop system except that it is done but once rather than at the start of each line.

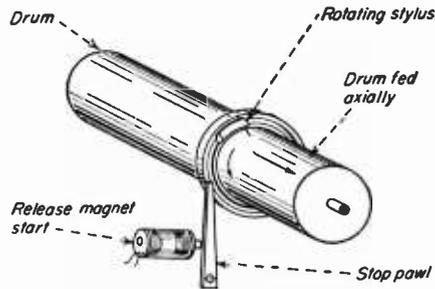


FIG. 17. One form of start-stop synchronizing mechanism.

PROPAGATION

The communication medium to be utilized has a great bearing on the design of the facsimile system. There are essentially three mediums used at present:

1. Radio
2. Submarine cable
3. Landline telephone circuits

Each has its special problems of propagation.

16. Radio Circuits. Long-distance short-wave circuits are affected by multipath reflections from the ionosphere which result in both general and selective fading variations of short time, diurnal, seasonal duration, and in other phenomena. They are also affected by the signal-to-noise ratio that can be realized with the receiving equipment now available, as well as by interferences from natural and man-made sources.

Short-distance v-h-f and u-h-f circuits will greatly minimize these difficulties provided they are limited essentially to line-of-sight paths and are quite sharply beamed. For broadcast work, even these circuits are affected by multipath transmission, in this instance, however, owing to reflections from buildings, hills, woods, airplanes, etc., which are somewhat to the side of the direct path between radio transmitter and receiver. Some of these reflection points and paths are stable and others are not, resulting in amplitude variations extraneous to the desired signal modulation.

Except where conditions are such that the v.h.f. or u.h.f. is reliably stable in amplitude, the use of f.m. in some form greatly minimizes these difficulties. The f-m principle can be applied as (1) a frequency variation of the r-f carrier with picture signal, and subsequent limiting of r.f. or i.f. in the receiver before the discriminator, or (2) a frequency variation of the audio subcarrier and subsequent limiting thereof in the receiver audio circuits before discriminator and rectification for recording, or (3) a combination

of subcarrier f.m. superposed as the audio modulation on a normal f-m radio transmitter-receiver system.

So long as the received signal amplitude is greater than the peak noise, the noise has but a minor effect. If the signal is two or more times the peak noise, the full benefit of the "noise-improvement threshold" is realized and essentially no noise appears in the recording. The same is true of interfering signals. Further, the only effect of the multipath phenomena occurs at the edges of sudden changes in the picture tone value, and this appears merely as a raggedness in the recording of such edges—no effect of multipath is to be seen in areas of constant or slowly changing tonal values.

17. Submarine Cable. The attenuation characteristics of even the loaded submarine telegraph cables precludes the useful transmission of frequencies much higher than 100 cycles. For facsimile transmissions the d-c variations in the phototube are d-c amplified and applied directly to the cable rather than as a modulation on a subcarrier as is done in other systems. Therefore the signals are subject to earth currents produced by magnetic variations and are greatly affected by magnetic storms. Correction must be applied to offset the "zero wander"; this can be done successfully for slow variations but becomes more difficult when the rate of these variations approaches the pertinent frequencies of the facsimile signals, as may happen. Earth currents of 9 or 10 volts varying at a rate of 10 cycles, or currents of 50 volts varying at a much slower rate, can be compensated. The transatlantic cable is comprised of two sections, each of which has an attenuation droop of 30 to 90 db at the higher frequencies. These sections must be individually equalized by "signal-shaping" networks. It is also essential carefully to correct a considerable phase distortion existing on the cables. Operations are effected successfully when the total noise and earth currents are 5 per cent or less of the signal swing from black to white.

18. Wire-line Telephone Circuits. Extensive networks utilizing wire lines are in operation in the United States and throughout Europe, as well as in other parts of the world. These all use existing telephone channels and systems and rely on transmitting an a-f subcarrier, amplitude-modulated by the picture signals. Both double and single side-band methods are used. In the latter the upper side band is suppressed, the carrier set near the top of the telephone channel (say, 2,400 cycles) and the lower side band, extending down to approximately 1,200 or 1,000 cycles, is transmitted.

For circuits of any considerable length it is important to equalize the lines, both for amplitude and for phase distortions. It is essential that the over-all phase characteristic for the pertinent frequency band be nearly linear. Relatively small deviations will delay certain frequencies with respect to others and may produce all sorts of weird effects in the recording. While this is completely true of high-definition photographic recording, the inherent low definition of the direct visual recording methods permits much wider latitude in the phase characteristic requirements of the lines.

For high-definition photographic work, the over-all gain must also be held within extremely close limits as a variation of 0.1 db can be perceived by the eye and variations greater than 1 db are disagreeable. For visual recording, these figures can be relaxed by 2 or 3 to 1.

Some systems are equipped with networks comprised of special circuits which have been carefully equalized and adjusted and are used only for facsimile work. These may be extended by using ordinary telephone lines to interconnect them with other locations, usually on an emergency basis, for a sudden news event. Other systems utilize ordinary telephone toll facilities and take their chances on the quality and stability of the circuit. Some connect their equipment directly to the lines through repeat coils, and others connect by inductance coils coupled to the ringing box and coil of the ordinary telephone subscriber's station. Most of these networks are set up primarily for the handling of news pictures and are therefore designed for the utmost flexibility so as to meet any emergency of news occurrences.

Reflections of the signals on the line, caused by mismatched impedance points along the circuit, can cause trouble by "ghosting" or producing signals lower in amplitude and delayed in time, which, however, often are strong enough to record.

TAPE-FACSIMILE SYSTEM

Tape equipment is designed solely for message communication as opposed to picture or news matter. It produces a record on a narrow tape, much as do the better known telegraph printers. The method of recording used to date is that of a rapidly rotating spiral and an axial bar moved by a loud-speaker magnet in accordance with the signal. The recording has been done either with a carbon paper tape or by applying ink to the surface of the spiral through the medium of a felt roller saturated with the ink. The scanning lines are crosswise of the tape and are made at a rate of about 60 per sec. The tape is slowly advanced lengthwise so the texture of the lines is about 60 to 100 lines per in.

Two distinct methods of transmission have been developed. In the United States much work has been done to develop a phototube scanner along the general principles outlined in Sec. 4. This method actually transmits a facsimile copy of written or printed messages placed on a tape at the scanner. It is being developed for mobile services, such as police and aircraft.

The second method utilizes a special instrument which comprises a large number of cams, one for each character (figure, letter, or punctuation mark) to be sent. The cut of the cam is such that a contact operated thereby will send out telegraphic type mark and space signals, which, when recorded as above described, will form the shape of the desired character. Two designs of this instrument have been developed. In one the message is perforated in paper tape as though it were to be sent on a standard printing telegraph circuit. The permutations of the holes in this tape consecutively select and release the proper cams, as the tape is fed through the instrument.

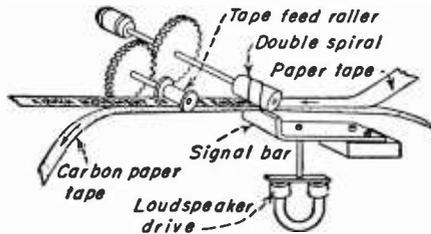


FIG. 18. Elements of facsimile tape recorder with double spiral.

In the other type a typewriter keyboard is manually operated and the striking of a type key will release the proper cam. This type of scanning is used extensively in Europe, and the United States rights have been acquired by one of the large companies in this country.

The synchronizing problem is just as pertinent as in the other systems but is possibly slightly easier. This is because the scanning line is so short and the rate so high that the discontinuity between the end of one line and the commencement of the next readily provides a frequency component that may be used for automatic framing and synchronizing.

TRENDS

The present trends of development and application of facsimile are toward higher speed, higher definition performance, and more automatic operation of the equipment. Applications are being found in broadcasting to homes on the 88- to 108-Mc band, in mobile and vehicular use such as aircraft, police and fire systems, forest service, in commercial communications, within large plants or buildings and between various plants and offices, in military installations of different types and widely varying requirements, as well as in point-to-point public service.

Speeds are being pushed upward of 100 to 200 sq in. per min and higher. Definition of 300 lines per inch has been provided with photographic recording, and the nonphotographic or so-called "visual" methods of recording are now being pushed toward the same goal. Circuit controls and operating adjustments are being minimized and circuit stability improved. Advanced design encompasses automatic loading and ejection features for the subject matter, automatic cutoff and ejection for the received

copy, remote starting and phasing of recorders so that they may be used for unattended operation, network installations with selective calling, break-in signals, etc

OPERATING STANDARDS

The Facsimile Committee (TR11) of the RMA has recently proposed the following standards for facsimile broadcasting in the 88- to 108-Mc band:

1. *Useful Scanning Line.* Shall be either 8.2 or 4.1 in.
2. *Line Advance.* Shall be 105 lines per in.
3. *Scanning Speed.* Shall be 360 lines per min. (Drum speed, 360 r.p.m.)
4. *Type of Modulation.* Standard facsimile broadcasting in the 88- to 108-Mc band shall be based upon the use of subcarrier a.m. until there have been demonstrated new developments desirable for adoption as new standards.
5. *Polarity of Modulation.* On subcarrier a.m. the maximum signal shall be radiated for black.
6. *Phasing Signal.* The nonpicture signal portion of the line stroke cycle (total scanning line) shall be 45 deg or $\frac{1}{8}$ revolution. The first 15 deg of this nonpicture signal interval shall represent transmission at approximately full white, the second 15-deg transmission at approximately full black, and the third 15-deg transmission at approximately full white, the synchronizing signal thus produced constituting a protected pedestal.
7. *Modulation Characteristic.* The subcarrier modulation shall normally vary approximately linearly with the optical density range of the subject copy (as the log of the reciprocal of the intensity of reflected light).
8. *Permissible Noise Level.* The subcarrier noise level shall be capable of being maintained at least 30 db below the maximum (black) picture modulation level.
9. *End-of-page Signal.* There shall be allotted a maximum of 12 sec for the transmission station identification, phasing or framing signals, remote control, page cutoff, or other special-purpose signals, immediately prior to the commencement of each page of copy.
10. *Copy Delivery.* Copy shall be delivered at the top in order that the recording will be done and the copy will be readable in a "straight-up" position.

The present standards for international transmission of facsimile, as established by the CCIR, are as follows:

1. Drum diameter, 88.00 mm (3.464 in.)
2. Drum circumference, 276.46 mm (10.88 in.)
3. Gripping (framing) loss, 15.00 mm (0.59 in.)
4. Phasing loss, 5.00 mm (0.196 in.)
5. Maximum skew or hunt, 0.08 deg
6. Drum length, 310.00 mm (12.2 in.)
7. Picture size, max, 250 × 290 mm (9.8 × 11.4 in.)
8. Drum speed, 60 rpm
9. Line advance, 4, 5 per mm (101.6, 127 l.p.i.)
10. International index of cooperation, 352, 440
11. Speed stability, 0.001 per cent
12. Frequency shift limits, 1,600 to 2,000 cycles
13. Standard frequency, 300 cycles or multiples

The index of cooperation as established by the IRE is given by the formula

$$M = CF$$

where C = circumference, or total length, of scanning line

F = fineness of scanning expressed in number of lines per unit length of drum axis

The international index is $\frac{1}{\pi}$ times the IRE index.

If two machines have different dimensions but the same index, the picture sent between them will be enlarged or reduced but will not be distorted in its proportions.

APPLICATIONS OF FACSIMILE

Radio Circuits. *Short Wave.* At relatively low speeds of 20 to 60 rpm for long-distance transmission.

U.H.F. At high speeds of 240 to 600 rpm for point-to-point work.

Medium Wave. At 360 rpm for broadcasting to homes.

Marine Service. Broadcast of weather maps, etc., to ships at 20 rpm.

Wire Line.

Point to Point. For news dissemination or public service, 90 to 120 rpm; for message pickup and delivery (customers' machines or "letter box" machines) at 180 rpm; for message service on trunk lines at 180 rpm.

Submarine Cable. At 20 rpm.

Photoengraving. Used for preparation of printing plates, either black-and-white or four-color separation plates for color printing.

Military. For both Army and Navy use in handling maps and documents.

Tape Facsimile or "*Hellschreiber.*" Used extensively throughout Continental Europe for news dissemination to agencies, by radio on 60 to 150 kc—also proposed for aircraft and police-car use.

References

Although a vast amount of work has been done in facsimile, it is fortunate that recent compilations have gathered the various references together so that the few citations given below will permit the reader to follow in detail the developments of facsimile in the radio field and will give him a working knowledge of the wire-line services.

Callahan, J. L.: A Narrative Bibliography of Radio Facsimile, in "Radio Facsimile," RCA Institutes Technical Press, 1938.

Jolley, L. B. W.: "Alternating Current Rectification," Wiley, 1928.

Mathes, R. E., and J. N. Whitaker: Radio Facsimile by Subcarrier Frequency Modulation, *RCA Rev.*, October, 1939, pp. 131-153.

"Radio Facsimile," RCA Institutes Technical Press, 1938. A group of papers detailing various aspects of the art.

Schroeter, Fritz: "Handbuch der Bildtelegraphic und des Fernsehens," Springer, 1932.

Wilson, J. C.: "Television Engineering," Pitman, London, 1937.

CHAPTER 21

RADIO AIDS TO AVIATION

BY HARRY DIAMOND¹

1. Importance and Trends of Radio Applications in Aviation. The success of any transportation system depends in a large measure upon the rigorous maintenance of safe, scheduled operation. Following the First World War, when the airplane became a potentially important factor in transportation, it was recognized that radio would prove a powerful tool in furthering an aviation transport industry. Today, with scheduled domestic air-passenger traffic averaging hundreds of millions of miles annually (294 million passenger revenue miles in 1946), with air mail and air express of proportional volumes, and with the range of air operations reaching around the earth, there is an increasing appreciation of the unique reliance which aviation must place upon radio aids.

Radio provides aviation with *communication, navigation, traffic-control, and collision-warning* services. From the time that the first piece of radio equipment was installed on a transport airplane, the ultimate goal has been the unification of these four basic functions into an integrated system directed to the continuous safe and efficient movement of air traffic. As the operational problems of air transport have become progressively more complex, there has been a gradual shift of responsibility for aircraft control from the individual pilot to the ground controller, and a corresponding shift in the detailed application of the four basic radio services. This transition is still in progress and is responsible, in large part, for the present fluid state of the art.

When the range of operations and the speed of the airplane were relatively low, and the volume of traffic was small and along a limited number of well-defined narrow airspaces (called *airways*), the principal aim of the radio aids was to furnish information to the pilot. Radio provided him with appropriate weather information, informed him when he was on or off his course along an airway, and marked out approach paths so that he could land even under conditions of poor visibility. The pilot was virtually in complete control of his flight plan. As the speed, range, and volume of operation of aircraft have increased, the airspace over populated areas such as the United States has become filled with an extremely complex network of airways connecting the locations between which air operations are conducted. Short-range traffic adheres to these narrow airways, whereas nonstop cross-country flights may traverse from airway to airway. The concept of the narrow airspace type of airway has given way to that of a universal airway² having an infinite number of navigational courses. The aircrew can no longer be provided with sufficient information to permit it to integrate its flight with that of other aircraft into an orderly flow of traffic. The problem of traffic con-

¹ Chief, Ordnance Development Division, National Bureau of Standards. EDITOR'S NOTE. The manuscript for this chapter was completed shortly before Mr. Diamond's death, June 21, 1948.

Wherever possible, direct acknowledgment for source of illustrations and text has been made in the form of footnotes and references. It has not been possible to make adequate acknowledgment to the Government agencies such as the CAA, U.S. Coast Guard, Air Matériel Command, Bureau of Aeronautics, Signal Corps, and others from which considerable help was received. Appreciation is also due to Aeronautic Radio, Inc., for assistance in the description of current operating procedures and to the commercial companies whose products are described. Much of the material presented is based on as yet unpublished papers so that the references are unfortunately not complete as to the publication source.

² For the purposes of traffic control, this universal airway is divided into areas of limited dimensions under the control of traffic centers, with a further division into still smaller areas centering on the airports and representing the focal points of traffic concentration.

trol has thus placed an overriding requirement on the organization of the basic radio services.

In considering the division of responsibility between ground controllers and crews of aircraft, the ground controllers must establish the flight plan to be followed and must make frequent checks of the degree of adherence to the prescribed plan. In the present state of the art, the aircrew must be given sufficient information to carry out the prescribed plan with minimum reference to the ground controllers. Ultimately, it appears necessary to by-pass the human navigator and to apply the navigational intelligence directly to the controls of the aircraft. In such a system, the aircrew should have means for correcting the flight of the craft if it does not adhere to the controlled plan or for deviating from the plan in case of emergency. The ground controllers should have continuous automatic monitoring of the position of each craft in flight with means for issuing control signals to conform to existing traffic requirements and flight conditions. Automatic ground computers and associated automatic pictorial display of the monitored data will form an important part of the control system. Although such a system is obviously not conceivable for the immediate future, it is presented for consideration as a background for the evaluation of the radio aids to be described.

Very large expenditures have already been made by governments and by the aviation industry in establishing extensive communication systems, flight-path beacons, landing aids, and schemes for traffic control. Responsible agencies must consider these present facilities from the point of view of possible continued utilization and expansion, rather than from abandonment in favor of newer systems. This fact has been given wide recognition in the long-term planning which is being conducted by the Special Radio Technical Division (COT) of the International Civil Aviation Organization (ICAO). This division, on which representation is world-wide, has made an extensive study of civil air-navigation requirements along international air routes and has prepared specific recommendations covering the evolution of a broad program of expansion and development.¹ The findings of its various committees represent the consensus of authorities on aircraft navigation and serve as a background for much of the material presented herein.

The initial expansion of aircraft transportation occurred during a period when radio communication by voice or telegraph, and radio directional equipment which utilized aural indications, constituted attractively simple facilities adaptable to aircraft control; but it was early recognized that eventually systems would have to provide many automatic features capable of covering every phase of flight. Pulse transmission and cathode-ray presentation were then unknown or had reached only the early stages of development. Directional aids utilizing visual indication on the aircraft and capable of integration into semiautomatic or fully automatic control had, however, been worked out well before the Second World War, both for use on a universal airway (the omnidirectional beacon and the automatic aircraft direction finder) and for landing aids (ILS). With the coming of the war, other types of navigational aids utilizing new techniques were developed. Certain of these aids proved of such extraordinary value in both offensive and defensive operations that many have since come to be considered indispensable future aids to civilian operations. Because these recent developments rely largely upon electronic equipment, much of which is expected in time to become fully automatic in operation, the need for communication facilities as a direct navigational aid should gradually decrease.²

2. Recognized Standards for Aviation Radio Facilities. Basic functional classifications of aviation radio facilities currently in operation or in an advanced state of development include (1) communication systems, (2) long-distance navigational aids,

¹ The recommendations of this international committee are directed only to international air routes. The 10 regions represented on the committee may individually adopt any preferred system of navigational aids provided the international air routes and terminals traversing the region are equipped according to the international standards.

² The present treatment of navigational aids to aviation will omit detailed discussions of communication facilities and equipment in accordance with this point of view.

(3) short-range navigational aids, (4) aids to final approach and landing, (5) airways traffic-control systems, (6) airport surveillance and control systems, (7) collision-warning systems, and (8) miscellaneous aids, such as distance-measuring equipment, location markers, and absolute altimeter.

A complete system for aircraft control must integrate the various facilities to afford continuous safe and efficient movement of air traffic (founded on effective navigational aids) along all portions of the routes to be flown, including take-off and landing. A number of integrated systems have been proposed and the more important ones will be described. Particular emphasis will be given to the system being implemented by the CAA to meet the special requirements of American civil aviation, as well as its obligations on an international basis.

Present recognized standards for aviation radio facilities are those formulated by the Special Radio Technical Committee of ICAO. The standards consist of (1) functional requirements as now envisaged, the satisfaction of which is an ultimate objective, and (2) immediate objectives outlining the requirements which can be met currently by services available or in an advanced state of development. The committee also published a set of criteria which it adopted for use in evaluating air navigational aids. The standard functional requirements and the factors for evaluating navigational aids are reproduced in the following two sections (in slightly abridged form) as of basic importance in the study of the material presented in this chapter.

1. Functional Requirements.

General. The communication portion of the basic system shall provide for the immediate exchange of intelligence over all portions of the route to be flown. The radio navigation portion of the basic system shall make it possible for the aircraft to be navigated over all portions of the route to be flown, including take-off and landing, without visual reference to the ground, under all flying conditions, and without assistance from a Ground Controller. The air traffic portion of the basic system shall provide for the safe, orderly, and expeditious flow of air traffic throughout its entire movement. The collision-warning system shall provide a safe and positive means for the avoidance of collision with other aircraft or obstacles. Air-horne equipment shall impose the minimum detriment to aircraft performance. The system shall be capable of use in all parts of the world.

Radio aids to final approach and landing. The navigational aids provided for final approach and landing shall make it possible safely to fly an aircraft on a predetermined approach path to a landing or a lighting on water terminated by a normal stop. In the final approach path and landing area, air ground communication shall be constantly available and free from static. Radio aids supplemented by adequate runway and approach lighting are basic requirements. The system shall be readily identifiable as to aerodrome and runway or water-channel served.

The system shall provide to the pilot precise, continuous, visual indication of the displacement of the aircraft relative to the predetermined approach path; precise, continuous, visual indication of the distance of the aircraft from a specified point in the landing area; immediate and positive indication of malfunctioning of any part of the system. The system shall be capable of providing for either automatic or manual flight to a landing.

The system shall not preclude simultaneous final approaches and landings on multiple runways or multiple water-channels; shall not be the limiting factor in the determination of the traffic handling capacity of the final approach zone and aerodrome; and shall be capable of furnishing reduced service to aircraft having limited or partially inoperative equipment. The navigation and communication portions of the system shall be capable of simultaneous operation.

Short-distance aids to air navigation. The short-distance navigational aid system shall present to the pilot in a convenient and natural form the information required, in a manner directly applicable to the handling of the aircraft. Such information shall be capable of presentation in other parts of the aircraft.

The system shall provide to the pilot continuous visual indication of distance and azimuth from any geographical point which he may select within the coverage area of the system and provide him with continuous visual indications which will allow him to follow any operationally desirable track and help him toward making good a previously declared Estimated Time of Arrival (ETA); shall enable the pilot to determine at any time his geographical position by a recognized method of fixing relative to the ground; shall be capable of forming an integral part of an over-all system providing automatic take-off.

en-route flight, position reporting, air-traffic control, and landing; shall not limit the utilization of aerodromes and airspaces; shall be free from disturbances in wave propagation and provide satisfactory service in all conditions of weather and visibility; shall provide immediate and positive indication to the pilot of the aircraft of the malfunctioning of any part of the system; and shall be capable of furnishing reduced service to aircraft with limited or partially inoperative equipment.

Aids to air-traffic control shall cover area, approach, and aerodrome control. Such aids shall provide for the expeditious control of the movement of aircraft in the air and on the ground; the safe and orderly routing of aircraft into designated flight paths; the separation of aircraft in the airspace and on the ground; and the utilization of available runways to their maximum capacity.

The system shall provide immediate, reliable, and interference-free communication between the aircraft and air-traffic control and between ground controls; visual presentation to the pilot and to air-traffic control of instructions or information to and from the aircraft; complete air-traffic control operated either automatically or manually on the ground and with either automatic or manual flight of the aircraft; freedom from language difficulties; continuous and instantaneous presentation of the traffic pattern, as required by air-traffic control, of the plan position, altitude, and identity of aircraft; immediate and automatic warning of danger of collision within the control zone; immediate and positive indication of malfunctioning of the system, with provision for the continuation of safe flight in the event of such malfunctioning; information concerning the nature and location of turbulent or hazardous weather; and shall furnish reduced but safe service to aircraft having limited or partially inoperative equipment. The system shall be capable of forming an integral part of an over-all system providing automatic take-off, en-route flight, and landing.

For approach-zone control, the system shall additionally provide for the expeditious and safe control of all traffic, including all facilities necessary for furnishing immediate instructions and guidance to such traffic; shall provide surveillance of aircraft in the approach zone; shall provide for the selection of aircraft intending to land and for guiding them into the final approach path to achieve a regular and rapid rate of landing; shall be capable of providing for the holding or delaying of aircraft; shall provide accurate monitoring of aircraft approaching, landing and taking off; and shall be capable of controlling the movements of aircraft during approach and during a landing or alighting on water, terminated by a normal stop.

For movement-area control, the system shall additionally provide for the safe and expeditious movement of all traffic in the movement area, including all facilities necessary for furnishing immediate instructions and guidance to all such traffic; shall be capable of use in all weather, visibility, and surface conditions; shall be capable of directing the movement of traffic in the movement area; and shall provide adequate separation of aircraft and/or vehicles, among themselves and from other obstructions.

Long-distance aids to air navigation. The long-distance navigational aid system shall present within the aircraft in a convenient manner all the information required; information directly applicable to the manipulation of the aircraft shall be capable of presentation to the pilot.

The system shall make continuously available an indication of the aircraft's geographical position by a recognized method of fixing relative to the ground; shall provide to the pilot continuous visual indications which will enable him to follow any operationally desirable track; shall be capable of satisfactory operation despite atmospheric or propagational disturbances and the effects of terrain; shall be capable of being integrated into an over-all system providing automatic take-off, en-route flight, position reporting, air-traffic control, and landing; and shall provide immediate and positive indication in the aircraft of malfunctioning of any part of the system.

The system shall be capable of simultaneous use by an unlimited number of aircraft. The navigation and communications portions of the system shall be capable of simultaneous operation. Indication of absolute altitude above the earth's surface shall be provided in the aircraft.

The ground equipment of the systems [mentioned above] shall automatically maintain its proper alignment; shall automatically and radically change the character of its indication in case a divergence from proper alignment remains outside the tolerance for the particular installation; and shall provide immediate and positive indication of malfunctioning to the appropriate ground authority. The navigation and communication portions of the systems shall be capable of simultaneous operation.

Radio aids to collision warning. The fundamental means for ensuring the avoidance of collisions should be an efficient air-traffic control system.

The basic system shall continuously provide safe and positive means for the avoidance of turbulence and of dangerous clouds, and of collision with aircraft and with obstacles; shall provide means of distinguishing between types of hazard, that is, shall be capable of conveying to the pilot information as to whether the object constituting a hazard be turbulence, cloud, aircraft, or other obstacle; shall not depend upon alertness of, or calculation or manipulation by, any member of the crew of an aircraft; shall give indication of the action required to avoid the hazard, such indication to be given in sufficient time to permit the pilot to take the necessary action; shall provide immediate and positive indication to the pilot of malfunctioning of any part of the system; shall be capable of discriminating between immediate and future hazards to preclude the commencement of avoiding action before such action becomes necessary; and shall be capable of operation simultaneously with the communication and navigation portions of the air-borne equipment.

2. Factors in Evaluation of Proposed Navigational Aids.

Operational adequacy. Accuracy; coverage or range (area or line, as appropriate); ability to provide immediate indication of failure or improper functioning; speed of interpretation for application; ease and accuracy in converting presentation to flight application (flyability); ability to meet operational requirements other than those specified in this paragraph.

Reliability. Susceptibility to partial or complete failure, to natural or man-made interference, to error or fading due to propagation characteristics or terrain, to human error, to saturation, to ambiguous indications; stability of courses or tracks.

Operational convenience. Simplicity and ease of maintenance of manipulation of air-borne and of ground equipment; suitability of form of presentation for the user and for direct use by the pilot or with the automatic pilot; freedom from language difficulties.

Adaptability to air-traffic control requirements. Minimum lateral and longitudinal separation of aircraft which may be safely permitted by use of the aid; capacity of the aid to position and guide aircraft for the purpose of achieving a high rate of landing; ability to make provision for aircraft which experience failures in flight or which are partially equipped.

Versatility, adaptability, and future potentiality. Capability of progressive evolution to meet future needs; adaptability to integration into any present or proposed systems of communication or navigational aids; suitability for search and rescue purposes; adaptability to meet the basic requirements of high or low density areas, and ability to expand; ability to provide auxiliary service if primary service fails; suitability for warning against collision and dangerous turbulence; suitability and convenience for installation in all types of aircraft; and adaptability to utilization with water aerodromes.

Frequency considerations. Band width required and position in spectrum; total required assignment of frequency spectrum for extensive utilization of the aid; economy of use of the spectrum in relationship to the service given; flexibility of equipment in respect to a change of frequency allocation, as might become necessary; effect on operational characteristics of the aid of such a change of frequency; and possibility of causing interference to other services.

Siting. Technical and administrative siting requirements.

Availability of equipment. Sources of supply and estimated date of availability in tested form.

Costs. Initial installed, operating, maintenance, and training costs; weight, bulk, power, and antenna requirements for airborne equipment; value to be derived from ability to perform additional functions.

3. Organization of Civil Aviation Radio Facilities of the United States. Civil aviation radio facilities can be broadly classified according to who furnishes the services: government systems and transport company systems. The government system is operated for the benefit of all fliers; transport companies plan their systems to incorporate the various services rendered by the government. Since 1920, when the Post Office Department established a series of aeronautical radio stations in connection with the operation of the air-mail service, the Federal government has steadily widened its participation in the field of air navigation and air-traffic control until it owns and operates hundreds of radio stations and other aids requiring the services of thousands of personnel.

The CAA is the government organization responsible for the installation, maintenance, and operation of civilian aviation radio facilities, for research and development leading to their improvement, and for the dissemination of information. Installation

and maintenance functions, which include the planning, design, construction, installation, and improvement of air navigational aids, are carried on by the Air Navigation Facilities Service; the actual operation of facilities is carried out by the Airways Operations Service; and research and development activities are conducted by the Technical Development Service.

Facilities directly operated by the CAA as of July 1, 1947, are listed in Table 1; data indicating the extent of utilization of these facilities are also presented in this table. The Federal airways system within continental United States is shown in Fig. 1.

Table 1. Extent of CAA Facilities for Air Navigation and Control

1. Number of facilities, as of July 1, 1947:	
a. Mileage of civil airways.	
(1) Total.....	73,393
(2) Under traffic control.....	44,057
(3) Oceanic routes.....	14,400
b. Air navigation radio aids.	
(1) Adcock ranges (RA and MRA).....	255
(2) Loop ranges (RL, MRL, and ML).....	125
(3) V-h-f ranges (VAR and VOR).....	72
(4) Total radio ranges.....	432
(5) "H"-type markers (radio beacons on airways).....	78
(6) "K"-type markers (compass locators).....	42
(7) Fan markers (full-powered).....	236
(8) Fan markers (low-powered).....	28
(9) Instrument landing system (ILS).....	51
c. Radio communications stations.	
(1) Interstate.....	398
(2) Overseas-foreign.....	8
d. Traffic control.	
(1) Airport traffic-control towers.....	124
(2) Airway traffic-control centers.....	29
e. Landline facilities (circuit mileage).	
(1) Weather-reporting circuits.....	60,273
(2) Traffic-control circuits.....	68,565
2. Number of operations, during fiscal year 1946:	
a. Conducted by airport traffic-control towers (including approximately 2,500,000 for the benefit of the Armed Forces).....	
	9,591,598
b. Conducted by airway traffic-control centers.	
(1) Aircraft operations (including approximately 6,800,000 for the benefit of the Armed Forces).....	10,294,691
(2) Fix postings (including approximately 2,140,000 for the benefit of the Armed Forces).....	8,157,272

Other agencies of the government have duties requiring the operation of radio facilities which can be utilized by civil aviation. Among these are the Army, Navy, Coast Guard, and FCC. As a service to pilots, the CAA maintains close liaison with these facilities, disseminates up-to-date information on their status or on sources of information regarding their status, and, where needed, establishes services to facilitate their use.

Current information on air navigation radio aids is given in the following publications:

Flight Information Manual. U.S. Department of Commerce, CAA, Office of Aviation Information. Issued semiannually. Includes air-traffic control and communications procedures, instrument approach procedures, direction-finding station data, airport and radio facility indexes, foreign entry and flight requirements, and radio call letters and frequencies.

Airman's Guide. U.S. Department of Commerce, CAA, Office of Civil Aviation. Issued biweekly. Includes information on new facilities, facility changes, notice to airmen on temporary conditions, such as airway construction or inoperability of facilities.

Radio Facility Charts. U.S. Department of Commerce Coast and Geodetic Survey. Revised weekly. Contains 42 separate charts providing complete radio information to facilitate the planning and execution of cross-country flights in the U.S. Shows all radio facility data with detailed tabulations of all radio stations in operation, classified as to aids to navigation and control towers.

Catalog of Loran Charts and Service Areas. U.S. Navy Department Hydrographic Office. *H.O. Pub.* 1-L. Includes description of loran system, discussion of factors affecting accuracy, shows ground-wave and sky-wave coverage, and presents a chart index of loran service in the Atlantic and Pacific Oceans.

Army Air Forces Radio Facility Charts. Headquarters, Air Materiel Command, Wright Field, Dayton, Ohio, Attention TSM DA2R Tech. Order 08-15-1. Issued monthly. Furnishes data and charts for use by air-force pilots; shows how military aircraft utilize virtually the same radio navigational facilities as do the civil aircraft.

U.S. GOVERNMENT FACILITIES

These comprise an integrated system of many diverse types of installation. The following brief descriptions of these facilities are intended to explain their functions; technical details will be presented in subsequent sections. A brief outline of future CAA plans for improvement of the integrated system is included in the present section.

4. **Radio-range stations** provide radio-marked courses for delineation of lines of flight during restricted visibility conditions and over-the-top operations when operations by visual contact with the ground are impossible.

The first type applied to the airways was the *l-f directional, four-course, aural* radio range. Some 360 are still in use. These operate generally at radio frequencies within the band 200 to 400 kc, with a modulation frequency of 1,020 cps. Each range has four courses, produced by the alternate interlocked keying of two figure-of-eight patterns, one "N" and the other "A", which are directed at approximately right angles to each other, the exact direction of each course being that which will best serve the airway and/or airport. Ranges are normally located within 2 to 4 miles of an airport, at least one course being aligned in such a manner as to facilitate instrument approach to same.

The position of a course is determined by comparing the relative strength of the signals N and A, the proximity of the course being indicated by approaching equality of these two signals. Exactly on course the interlocking of the two signals produces a steady monotone (1,020 cps), interrupted only for the station identification. The latter, consisting of a two- or three-letter combination, is transmitted twice every 35 sec, first in the N quadrant and then in the A quadrant.

Voice transmissions are made on the range frequency unless otherwise indicated by the inclusion of the letter W in the station designation.

These l-f directional ranges can be further classified by type of radiator utilized and by power output as follows:

Tower Radiators. Five steel towers approximately 130 ft in height, four forming a square and one in the center, are utilized for the following designations:

RA, power output between 150 and 400 watts, usable distance 65 to 100 miles.

MRA, power output between 50 and 150 watts, usable distance 40 to 75 miles.

Loop Radiators. Two crossed, vertical loops mounted on five poles approximately 50 ft in height, four poles forming a square and one in the center, are utilized for the following designations:

RL, power output between 150 and 400 watts, usable distance 50 to 75 miles.

MRL, power output between 50 to 150 watts, usable distance 30 to 55 miles.

ML, power output less than 50 watts, usable distance 20 to 40 miles.

The *v-h-f directional radio range* was developed to introduce a number of operational advantages over the l-f type, to be discussed later. The designation applied to the v-h-f directional radio range is VAR. This is a two-course visual, two-course aural range with simultaneous voice feature operating within the band of 112 to 118 Mc. The power output is approximately 200 watts, and the usable distance is normally somewhat greater than line of sight, depending upon the altitude of the ship and the terrain, with an upper limit of about 250 miles at very high altitudes. The radiators employed are small and closely spaced, being located in a small building mounted on a tower 30 to 60 ft in height; a counterpoise approximately 35 ft in diameter is also mounted on the tower.

Visual Courses. The visual courses of VAR are 180 deg apart and are created by the overlapping of two field patterns essentially cardioid in shape, one modulated at 90 cps and the other at 150 cps. An electrical indicating (pointer) instrument, actuated by a receiver designed for the purpose, is utilized in the aircraft to locate and fly

the courses. This is accomplished by comparing the relative strengths of the two opposite sectors. The indicating pointer, or needle, is centered (vertically) on the instrument scale when the aircraft is exactly on course. Full-scale deflection of the needle will occur at approximately 10 deg off course.

Aural Courses. The aural courses of VAR are also 180 deg apart and are located at right angles to the visual courses. The aural courses are produced by the alternate interlocked keying of the letters N and A in a manner similar to that of the l-f directional range except that the field patterns are like those utilized for the visual courses of VAR. The modulation frequency of the aural signal is 1,020 cps and, like the l-f directional ranges, the aural signal is interrupted approximately twice a minute for transmission of the station identification. The orientation of the N and A sectors with respect to the visual courses is always the same; *i.e.*, N will be on the east leg of a visual east-west course or on the north leg of a visual north-south course.

The omnidirectional radio range radiates a signal which through the use of special receiving equipment may be utilized to establish a visual course in any radial direction with respect to the transmitting station. Such a station may be considered to have an indefinite number of flyable courses. It was developed to provide the type of service required by the concept of the universal airway. An "azimuth selector" dial enables the pilot to select a course along any direction toward or away from the range station, special means being provided to resolve ambiguity. The pilot follows the selected course by keeping the pointer of the indicating instrument centered (vertically).

The omnidirectional range radiates a nondirectional field modulated by a l.f. of constant (fixed) phase and a directional (figure-of-eight) field rotating in azimuth at the same l-f rate. The amplitude of the signal at a fixed point therefore varies sinusoidally with an absolute phase that depends upon the azimuth of the receiving point. This phase is automatically and continuously measured by the air-borne receiving equipment in terms of deviation from the reference phase to produce a direct indication of radial position or bearing.

This facility may be used for determining the true bearing of the aircraft with respect to the station by tuning in the station and adjusting the azimuth selector so that the instrument pointer centers, *i.e.*, reads on course. The selector dial then presents the desired bearing reading. An "azimuth indicator," operating independently of the selector dial and course indicator and given direct bearing information, has recently been developed. Bearings obtained from two or more stations will enable the pilot to determine his geographic position.

Omnidirectional ranges are classified by frequency as follows:

LOR, l.f., operating generally between 70 and several hundred kc and with a carrier power output of approximately 10 kw. In appearance the LOR station resembles the RA station, except that towers 320 ft high are used for the LOR. Usable distance will vary with terrain but will generally be in excess of 600 miles, atmospherics permitting.

VOR, operating within the band of 112 to 118 Mc. In appearance, power output, and usable distance the VOR station is similar to the VAR station.

The station identification of the VOR type is transmitted intermittently on the carrier, utilizing a 1,020-cps keyed tone, except during voice transmissions. Simultaneous voice transmissions are made on the range frequency of all VOR stations unless otherwise designated by inclusion of the letter W in the station-type designation. Voice transmissions are not contemplated from the LOR stations, nor is it planned to provide for identification keying.

The announced plans of the CAA call for the completion of the installation of v-h-f omnidirectional radio ranges (VOR) on all the existing airways now served by l-f directional ranges and for equipping all new airways with this type. There are in operation some 72 v-h-f ranges, certain of which are of the directional type and certain of the omnidirectional type. The directional-type ranges (VAR) were installed prior to completion of development of the omnidirectional type and are being converted to the latter type. The ultimate plan is to have a network of v-h-f omnidirectional beacons

(VOR) throughout the United States. In the interim, installations of the older range types will be kept in operation.

Present CAA studies indicate that a grid of 15 l-f omnidirectional beacons (LOR) installed at suitable locations in the United States will furnish complete coverage for cross-country flight operations within the United States and will also serve considerable areas in all directions outside the boundaries of the country. The tentative plan schedules the following areas where LOR stations should be installed: Seattle, Wash.; San Francisco, Calif.; San Diego, Calif.; Salt Lake City, Utah; Great Falls, Mont.; Cheyenne, Wyo.; Omaha, Neb.; El Paso, Tex.; Brownsville, Tex.; New Orleans, La.; Minneapolis, Minn.; Fort Wayne, Ind.; Nantucket, Mass.; New Bern, N. C.; and Miami, Fla.

Because of CAA responsibility for equipping Alaska with radio aids similar to those on the continental U.S. airways, and for equipping Hawaii and the Philippines to serve air traffic and to connect them with international and transoceanic routes, VOR and LOR installations are planned at appropriate locations. A sizable number of VOR installations will be required. It is anticipated that at least two LOR installations will be made in the near future in Alaska and at least three in the Pacific Ocean area.

5. Distance-measuring Equipment (DME). This device, an adaptation of wartime radar, will enable the pilot to determine at all times his distance from the transmitting station. The airborne equipment comprises a transmitter and receiver, the former unit transmitting short high-amplitude pulses approximately 200 times per second. These pulses are received at the ground station, which likewise comprises a transmitter and receiver, the receiver being tuned to the frequency of the air-borne transmitter. Each time a pulse is received on the ground, the associated transmitter is immediately triggered and sends a pulse reply. When this pulse from the ground is received by the plane, circuits in the plane equipment measure the time interval between the interrogating pulse originated by the plane and the reply received from the "transponder" on the ground. This time is converted directly to miles and indicated on a meter.

The air-borne equipment consists of a transmitter, receiver, indicator meter, and associated gear, the combined weight of which is approximately 25 lb. The ground equipment consists essentially of a receiver and transmitter. Current equipment of this type operates in a band centered at 1,000 Mc. A multifrequency channel system will be employed. A coding system for identification of each DME beacon is incorporated in the installation. In operation, a beacon is capable of replying to as many as 50 air-borne units at the same time.

The ideal in navigation of aircraft under instrument conditions is to obtain continuous direct readings of distance and direction from one or more ground points. The omnidirectional beacon provides a solution of the problem of direction; the solution of the problem of distance measuring can be achieved by the installation and use of DME. Such installation is planned by CAA at all v-h-f omnidirectional stations (VOR) and at all instrument landing installations (ILS). The system will provide flight paths especially convenient for accurate manual or automatic following: (1) along radial courses to or from the ground stations; (2) along circular orbits of any radius about these ground stations; and, (3), with an R/θ computer added, along any straight-line path offset with respect to the ground stations.

6. Loran. This is a long-range navigation system developed during the war and widely adopted throughout the world. It operates on the following principles: radio signals consisting of accurately synchronized short pulses (80 μ sec) are broadcast from a pair of special transmitting stations (a "master" and a "slave"), separated by from 200 to 400 miles, and operating on a common frequency within the band 1,800 to 2,000 kc. Each pair of stations comprises a loran facility or "rate." These signals are received aboard the ship or aircraft on a special receiver, and the difference in times of arrival is measured on a special indicator. This measured time difference is utilized to determine directly from special loran tables or charts a line of position which (neglecting the oblate curvature of the earth's surface) is a hyperbola with the two transmitting stations of the loran rate as foci. Two intersecting lines of position, determined from two properly selected loran rates, are required to obtain a loran fix.

A number of loran rates operate on the same radio frequency, but the pulse recurrence rate differs for each facility. By setting the receiving equipment for the proper frequency and recurrence rate, the navigator is assured of the identification of the facility utilized.

Loran provides accurate fixes, within 1 per cent of the distance range, to distance ranges over water of about 700 nautical miles by day and about 1,400 nautical miles by night. The daytime range over land is materially reduced by the attenuation of the ground wave and, in the r-f band indicated, would be of the order of 200 miles.

Loran service is now available over approximately 60 million square miles or three-tenths of the entire earth's surface, most of the service area being over water. The American stations are operated by the U.S. Coast Guard and represent 22 of the 26 loran facilities throughout the world. The geographical distribution of loran facilities follows: Atlantic Ocean, 10 rates; U.S. Pacific Coast, 4 rates; Hawaiian Islands, 2 rates; Marshall Islands, 2 rates; Marianas Islands, 2 rates; Iwo-Tokyo-Okinawa, 2 rates; Philippine Islands, 2 rates; Aleutian Islands, 2 rates. Navigation by loran service is thus feasible along both coasts of North America, along the great circle courses of the North Atlantic and North Pacific, and in the Central and Southwest Pacific areas.

Transoceanic airlines using loran are: American Overseas Airlines, B.O.A.C. (Atlantic Division), K.L.M. Royal Dutch Airlines, Pan American World Airways, and Trans Canada Airlines. As of Mar. 1, 1947, at least 21 "scheduled" extended overwater flights arriving at, or departing from, Honolulu made use of loran in the Pacific area.

The majority of loran receiver-indicator equipment now in use was developed and manufactured during the war and requires that time-difference determinations be made by manual manipulations involving the matching of the master and slave pulses and then switching to additional scope selection positions to count and extrapolate between time marker divisions. Thus, determination of a fix may require several minutes. Receiver-indicators now being manufactured and installed incorporate direct-reading counters, so that once the pulses are matched, a direct reading of time difference is obtained. Current developments indicate that it may prove feasible to secure automatic matching of the pulses, thereby further reducing the time for a fix.

L-f loran is an adaptation of loran at lower frequencies primarily to increase distance range of useful operation by reduction of attenuation of the ground wave (particularly over land) and by the attendant reduction of the errors introduced in loran by use of the sky wave. L-f loran has had successful operational service in central and northern Canada on a frequency of 180 kc, and experiments have been conducted on still lower frequencies. Useful distance ranges up to 1,000 miles over land and 1,500 miles over sea appear practicable. One method of pulse matching, called "cycle matching," affords unusual possibilities of accuracy and will be described in detail in a later section.

The short- and long-range navigational aids, described above, were evaluated, along with other available systems, in the technical discussions of ICAO from the viewpoint of standardization on international airways. It was concluded that no aids were sufficiently proven to justify universal adoption to the exclusion of competing aids and strong recommendations were adopted toward the continued development of promising types.

The Technical Division recognized, however, that a certain degree of uniformity of equipment was requisite along the important international routes and recommended the following:

1. Adoption of v-h-f omnidirectional ranges, together with DME, for operation until at least Jan. 1, 1955

2. Retention and extension of loran as necessary to meet traffic requirements

3. Installation of l-f loran to investigate coverage of critical areas within the North Atlantic, Southeast Asia, Africa, Australasia, Pacific Ocean, and South Atlantic, such installation to be limited to those necessary to meet the operating requirements on international air routes in these areas

7. **Marker Beacons.** Radio marker beacon stations of two essentially different classes, v-h-f and l-f, are located along the civil airways to supplement the radio-range system.

V-h-f markers operate on 75 Mc and include the following types:

Station location marker, designation Z. Serves to locate positively the range station with which it is identified. Nominal power output 5 watts, unkeyed, 3,000 cps modulation, vertical cone radiation pattern.

Fan marker, designation FM. Serves to furnish aircraft with a check as to progress along airways, to mark junction of two radio-range courses, to designate "hold-over" point along airport approach 10 to 20 miles from airport, or to mark obstruction. Nominal power output 100 watts, keyed for identification, 3,000 cps modulation, fan-shaped radiation pattern extending (at 1,000 ft altitude) 2 to 4 miles along the range course and 12 miles across the course.

Low-power versions of the fan marker, designation LFM. Nominal power output 5 watts or less; used for special purposes such as for "let-down" along a radio-range course. The

lower power is required for installations closer than 8 miles to an airport, to reduce interference with Z marker.

L-f markers are nondirectional radio beacons operating in the band 200 to 400 kc, and serving principally for use with automatic direction finders (ADF) aboard the aircraft. Station identification, utilizing a 1,020-cps keyed tone, is transmitted intermittently on the carrier. The several types are:

Type H. Power output 50 watts or greater; voice transmission feature unless designation is followed by letter W, operates continuously.

Type MH. Power output less than 50 watts; operates on request, usable distance range 25 miles.

Type K. Power output 25 watts; operates continuously, generally complements ILS boundary markers, and thus serves as a l-f landing aid. Serves as an interim aid, where ILS is not available, to facilitate instrument landing approaches utilizing one of the range courses. Also enables pilot to follow a more precise "holding" pattern during airport traffic-control operations.

8. Instrument Landing System (ILS). This was developed to enable properly equipped aircraft to navigate to a safe landing during poor weather conditions without visual reference to the ground. It serves also to reduce the length of time aircraft must be held aloft, during periods of congestion, by speeding up instrument approaches and landings.

A standard instrument landing system ordinarily consists of a localizer unit (LO), a glide-path unit (GP), and two position markers, each with a homing beacon, designated as "outer marker" (OM) and "middle marker" (MM). Under certain conditions, a third "boundary marker" (BM), without a homing beacon, may be required. In some cases, the homing beacon at the middle marker is omitted.

The *localizer unit* provides a range course for horizontal guidance. It is essentially a two-course v-h-f radio range aligned with the direction of approach. It operates in the band 108.3 to 110.3 Mc, with a maximum power output of 200 watts, and may be either of the "equisignal" or of the "phase-comparison" type. In either case, it actuates the vertical pointer of the cross-pointer "combined" instrument on the instrument panel of the aircraft, thereby providing continuous indication of the lateral position of the aircraft with respect to the approach course. With the phase-comparison localizer, identical receiving equipment may be used as for the v-h-f omnidirectional radio range. A simultaneous voice feature, for use by the airport controller, is provided on each type of localizer, together with station identification keying when voice transmissions are not being made.

The *glide-path unit* provides an equisignal course in the vertical plane containing the approach course and inclined at an angle of 2 to 4 deg with the horizontal as measured from the point of contact. It actuates the horizontal pointer on the aircraft "combined instrument" and provides continuous indication of the vertical position of the aircraft with respect to the desired "let-down" course. It operates in the band 332.6 to 335.0 Mc with a maximum power output of 20 watts.

The *marker beacons* are of the low-power v-h-f fan marker types operating on 75 Mc with power output of 2 watts. They are tone-modulated and identification-keyed, and may be identified aurally or by lights on the instrument panel of the aircraft. The fan-shaped pattern is directed vertically upward with the fan at right angles to the approach course. The outer marker modulation frequency is 400 cps and its identification signal consists of 2 dashes per second. The middle marker modulation frequency is 1,300 cps and it is keyed continuously with alternate dots and dashes. The boundary marker modulation frequency is 3,000 cps and it is keyed continuously at 6 dots per second. The three markers are located along the approach course, as follows: outer marker, 4.5 miles ($\pm 1,000$ ft), from the approach end of the runway and within 150 ft of the runway center line extended; middle marker, 3,500 ft (± 250 ft from the approach end) and within 50 ft of the runway center line; boundary marker, 250 ft (± 50 ft) from the approach end and 200 ft to the left of the runway center line (as viewed from the middle marker).

The *homing beacons* (type K) are nondirectional transmitting stations in the band 200 to 400 kc, with power output of 25 watts and operate continuously, unkeyed. They provide signals for automatic radio direction finders aboard the aircraft and thus furnish means for determining the heading of the aircraft with respect to the desired approach course.

Some 51 ILS installations were in operation in the United States as of July 1, 1947, with projected plans for additional installations at some 32 locations in the near future. CAA plans for the continuation of this installation program, depending on appropriations, until all scheduled air-mail and passenger stops, both in the United States and in the territories, have been implemented. It is estimated that approximately 165 installations are currently needed on U.S. airports; 25 of which at the more important terminals must be duplicated to serve two runways.

On the aircraft, the pilot maneuvers to pass over the outer marker at an altitude of about 1,000 ft, maintaining the aircraft "on-course" in both the horizontal and vertical planes by observing the intersection of the two pointers on the combined instrument. The aircraft passes over the airport boundary at about 50 ft altitude, and continues to a safe contact with the airport surface at a point allowing ample margin for a normal stop.

The value of ILS has been demonstrated in use of the installations in present operation. It was adopted as an ICAO standard for installation at all international terminals.¹ It was adopted, with some variations, as standard by both the Army and Navy during the war and became quite well known under the Army designation SCS-51. It has full potentialities for automatic landing of the aircraft, the signals operating the cross-pointer instrument being connected to an electronic coupling unit which controls the electric autopilot. Nevertheless, the problem of landing under poor visibility conditions is considered by responsible officials to be too important to place complete reliance upon ILS. Early in 1947, the Committee on Interstate and Foreign Commerce, U.S. Senate, recommended that the installation of ILS equipment be extended to 166 airports and that war-developed ground-controlled-approach (GCA) equipment be provided at 30 of the more important airports, the intent being that, in the present state of the art, it is necessary to use both systems at major airports as complementary aids rather than to place entire dependence upon either alone.

9. Ground-controlled-approach System. Developed by the armed services during the war, some 4 CAA, 7 Army, and 17 Navy GCA installations were available to civil aircraft as of July 1, 1947, either on a continuous or on an emergency (30-min notice) basis. The GCA system is a specially developed combination of search and precision scan radar units which permits its operators to "talk the airplane down" to the point of contact on the airport surface. Only the regular flight instruments and communications equipment are needed on the airplane.

In the wartime version, the GCA station is placed 500 ft to the side of the runway at the end opposite to which the airplane is to land. A 10-cm search radar scans a 35-mile area around the airport, and its operator directs pilots of inbound airplanes to the approach starting area at correct altitudes. This area is about 6 miles from the approach end of the runway.

Two precision 3-cm radar beams scan the approach sector in azimuth and elevation. These give to the ground operators precision measurements of distance and lateral and vertical displacement of the airplane. By proper instructions in voice to the pilot, the airplane is controlled by the ground operators to follow an approach path corresponding to a line drawn on the radar scope maps.

More recent versions of GCA, improved to conform more closely to civil aviation requirements, employ a three-dimensional scope with the information remotely indicated in the airport control tower. This eliminates the need for coordination between a number of operators and provides for control of all traffic by the traffic-control tower.

Because civil aircraft may not be able to receive the voice transmissions on Army and Navy assigned frequencies, an emergency v-h-f transmitting channel has been provided, on 140.58 Mc. The pilot desiring GCA service contacts the CAA airport traffic-control tower. The GCA voice transmissions are then made on the regular assigned station transmitting frequency, on 3,117.5 kc (for scheduled aircraft) or on 140.58 Mc. For aircraft not equipped to receive either h-f or v-h-f, the instructions may be relayed on CAA facilities in the 200- to 400-kc band.

10. Direction Finders. Radio direction finding facilities represent auxiliary aids to civil aviation available in two forms: the direction finder may be on the aircraft or it may be on the ground. For requisite accuracy with simple antenna equipment aboard the aircraft, direct space-wave reception is required; accordingly, low radio frequencies or very high radio frequencies must be employed. Since ground d-f installations may have receiving antennas of any required degree of complexity, there are no restrictions as to frequency bands. High frequencies in the range 2 to 30 Mc may be utilized with sky-wave distance ranges of hundreds, or even thousands, of miles giving average bearing errors of from 1 to 3 deg. Such distance ranges are particularly useful for transoceanic navigation. However, for overland operation, expediency has indicated the use of the direct space wave with corresponding distance ranges of the order of 100 miles.

Air-borne direction finders are utilized for homing purposes and for cross-bearing fixes. A four-band set, to cover the frequency range 100 to 1,750 kc, provides automatic bearing indication of the direction of arrival of r-f energy and simultaneous reception of intelligence.

¹ The equisignal localizer was specified for such terminals until 1951; CAA plans to use the phase-comparison type at other airports.

Such a unit may be utilized on any of the CAA 200- to 400-kc facilities with a distance range depending upon the power of the transmitting station, with the exception that the loop ranges (RL, MRL, ML) are not suitable for direction finder operations.

The U.S. Coast Guard operates nearly 200 radio beacons along the ocean coasts and the Great Lakes as a navigational aid for use by ship direction finders. These stations operating in the band 285 to 315 kc provide reliable average distance ranges (over water) of 200 to 400 miles from class A stations, 100 miles for class B, and 10 to 20 miles for class C and D. They utilize A-2 transmissions, are on 1 min and off 2 min to accommodate station grouping, and are identification-keyed during the 1-min period of transmission. At present, such keying interrupts the carrier wave. To facilitate the use of these stations by aircraft and marine automatic direction finders (ADF) the Coast Guard has converted 11 of the stations for continuous carrier transmission during the 1-min "on" period, with the keying applied as a modulation. More stations may be converted depending upon the degree of demand.

Another important facility that can be utilized by aircraft pilots for purposes of direction finding and fix determination consists of regular transmissions from commercial broadcast stations. All such stations having a power rating of 100 watts or more are listed in the "Flight Information Manual," both by operating frequency and by geographical position. Satisfactory use of broadcast station transmissions as navigational aids is dependent upon a proper understanding of the properties and peculiarities of radio direction finders, inasmuch as reflected sky waves often produce serious interference with ground-wave signals.

Ground-station direction finders take bearings on transmissions from the aircraft and, by triangulation between two or more stations, provide the aircraft with a position fix.

11. D-f Networks. Three radio d-f networks are available within the continental United States, operated by the FCC, by the Coast Guard, and by the Army, respectively. They are provided with direct telephone connections with the CAA. The FCC network is the largest of the three, comprising 10 primary d-f stations on a teletype hookup, and seven others in a radio network. All these stations can take bearings on any transmissions between 2 and 30 Mc. Fixes generally require the transmission of 20-sec dashes (either by key or by pressing the microphone button), interspersed with call letters, during a period of 5 min, after which a further period of 5 to 10 min is needed by the ground station for evaluation. The fix is then transmitted to the aircraft as readings of latitude and longitude and is classified as "good," "fair," or "poor," according to whether the range to the aircraft is less than 20, 40, or 60 miles, respectively. Aircraft pilots can obtain this d-f service merely by contacting any CAA communications station, any air-to-ground station operated by the Airways and Air Communications Service of the Army (AACCS), or any airport control tower. Although the operating frequency range is from 2 to 30 Mc, it is stipulated that, wherever possible, daylight transmissions from aircraft shall be between 6 to 8 Mc, and night transmissions between 4 and 6 Mc.

The Coast Guard operates three d-f stations within the United States, capable of taking bearings over a frequency range from 2 to 10 Mc, and also operates 11 communications stations maintaining continuous listening watches on the 8,280-kc emergency frequency. In utilizing this service the aircraft pilot is expected, if possible, to transmit suitable distress or urgent signals, along with such general calls as NCG or NCU, on either 500 or 8,280 kc, thus establishing contact with a Coast Guard station. Lacking the ability to transmit on one or other of these two specified frequencies, the aircraft must communicate with any CAA communications station, any air-to-ground AACCS station operated by the Army, or any airport control tower, and request telephone contact with the nearest Coast Guard district office and the subsequent relaying of instructions. Following transmission from the aircraft of 20-sec dashes and call letters, repeated the specified number of times, the Coast Guard station will furnish bearings and fixes in one of three different classes, according to the estimated accuracy of the readings. Bearing data are considered as class A, class B, and class C, for accuracies to within ± 2 deg, ± 5 deg, and more than ± 5 deg, respectively; and fixes are similarly classified as being accurate to within 5, 20, and 50 miles or more.

The Army operates a series of 21 v-h-f d-f stations which provide homing facilities for the benefit of aircraft operating within the continental United States. These stations have been located to conform with the density of air traffic, nearly one-half of them being located along or near the Atlantic seaboard and the remainder scattered throughout the rest of the country. They are operated by the AACCS for the benefit of all aircraft. Their reliable operating range depends upon the altitude of the aircraft and varies from about 55 miles at 1,000 ft to 250 miles at 30,000 ft. The stations transmit on two distinct fre-

quencies, which have been assigned to specific services, namely, 139.32 Mc, covering routine navigational homing for fighter aircraft; and 140.58 Mc, providing emergency homing for all aircraft. In utilizing these d-f facilities, the aircraft pilot contacts the v-h-f station directly, makes a short count by voice, from 1 to 5 and back, and receives instructions from the v-h-f station as to the azimuth to be flown.

12. CAA Communication Facilities. A CAA network of stations is located along the airways of continental United States and of U.S. territories, usually at airport or landing areas, with an average distance between them in the United States of about 80 miles. All radio range stations (RA, RL, MRA, MRL, and ML) all ILS localizers (LO), and all radiobeacon stations (H and MH) have voice transmission facilities. Radio range stations may interrupt the range signals for voice transmission on the same carrier frequency, may transmit the voice simultaneously on the same carrier frequency, or may broadcast both on the range frequency and on a separate frequency, usually VHF, 111.1 or 116.1 Mc. ILS localizers have simultaneous voice feature. Radio-beacon stations interrupt their identification signals for voice transmission, class H on a scheduled basis and class MH usually on request. All CAA communications stations equipped for voice communications with aircraft stand continuous listening watches on 3105, 4495, and 6210 kc, unless otherwise indicated in the "Airways Flight Manual." In addition, CAA communications stations maintain a continuous watch on 122.1 Mc as a special service to private fliers. The communication facilities are supplemented by extensive landline teletype and telephone circuits and by point-to-point radiotelegraph stations. Special overseas foreign aeronautical communication stations (OFACS), located along the boundaries of continental United States and in the U.S. territories or in foreign countries, afford equivalent service to overseas and intercontinental air-transport operations.

The continental system provides for two-way communication between the ground and aircraft in flight and for scheduled radiotelephone broadcasts of airway and meteorological information. The landline and point-to-point radio stations provide for the collection and dissemination (between stations) of weather reports and forecasts and of aircraft movements and control messages.

Weather broadcasts on the radio range and radio beacon frequencies take place twice an hour. At each station the material broadcast consists of the local weather report and the latest available surface weather reports from surrounding locations. In addition, a local winds-aloft report is broadcast four times daily. At selected stations the U.S. Weather Bureau provides a local terminal forecast covering the next 2 hr. This forecast is broadcast, when available, immediately following the local weather report. As an additional service to private fliers, "flight advisory" broadcasts are made concerning changes in weather conditions that occur after take-off, including reported icing conditions, suggestions as to alternate courses or altitudes, and other similar items of information.

In collecting weather information, the CAA has the cooperation of the Weather Bureau which maintains a large number of weather-observing stations at the airports and at points off the airways. By agreement with the Weather Bureau, the CAA installs and maintains all the teletype equipment at weather-observation and communication stations associated with the airways.

Air-traffic-control operations are integrated into the communication facilities setup. Control of air traffic along the airways is handled from air-route traffic-control centers and air traffic on and in the vicinity of airports is controlled from airport traffic-control towers. Control centers and towers are tied in to and utilize the voice transmission facilities of the nearest CAA stations. The control centers maintain a listening watch on 3,105, 4,495, and 6,210 kc for scheduled operations and on 122.1 Mc for private fliers. The control towers maintain a watch on the same three h-f channels and on 122.5 Mc for private fliers.

Three separate *teletype circuits*, a landline *interphone system*, and point-to-point radiotelegraph stations are required to implement the weather broadcast and air-traffic-control functions of the continental communications system as follows:

Service A. Collection and distribution by teletype and point-to-point radio of hourly and special surface meteorological reports, and of notices to airmen (NOTAMS) concerning field conditions, inoperative air navigation aids, etc.

Service B. Transmission and reception by teletype and point-to-point radio of aircraft movement information, including messages containing requests and approval to conduct an aircraft flight, flight plans, in-flight progress reports, aircraft arrival reports, etc.

Service C. Collection and distribution by teletype and point-to-point radio of synoptic meteorological reports; radiosonde and other upper air observations and weather-map data; airway, airway terminal and regional forecasts, weather advisories, etc.

Service F. Dissemination by interphone of messages designed to expedite the flow and to prevent the collision of aircraft under instrument flight rule conditions.

The overseas communication stations (OFACS) extend meteorological and control services to aircraft in flight along overseas or intercontinental air routes and in addition (by radiotelegraph or radioteletype transmission) exchange meteorological, traffic-control, and other information with similar stations located on U.S. territories and possessions and with foreign stations. This type of station comprises three units which are interconnected by land lines or radio: (1) the transmitter station having a number of separate channels suited to point-to-point long-distance operations and to ground-to-air operations. (Directional antennas are used for the former service with transmitter power ratings ranging from 8 to 20 kw; the antennas for the latter service are semidirectional or non-directional with transmitter power ratings from 350 watts to 3 kw); (2) the receiving station distant from the transmitting station to avoid interference and equipped to receive on appropriate frequencies with antennas corresponding to the service; and (3) the control station, generally located at or near an airport and equipped for point-to-point radiotelegraph and radioteletype transmission and reception and for radiotelegraph or voice contact with aircraft.

Service O. An auxiliary landline teletype network; interconnects the overseas communications stations located in the United States and provides for the collection and dissemination of national and international meteorological information and of aircraft movement information.

13. Radar Facilities. The adaptation of radar in aid to civil aviation is as yet in its early stages; accordingly, facilities in current operation are rudimentary and relate principally to surveillance functions in connection with traffic control. Nevertheless, analysis of the problems of controlling air traffic to prevent collision between aircraft and to expedite their movement, particularly under conditions of limited or no visibility, points to the future installation of extensive government facilities. The steady growth of civil aeronautics activities with the attendant increased congestion along principal air routes, particularly at airports, will emphasize the need and force the installation of: airport surveillance and precision beam radar for search and precision scan of aircraft in flight and for control of movement on the airport; radar beacons for use with air-borne radar to provide positive position fixes and distance measurements; storm-detection facilities; and other radar adaptations to perform functions not currently envisaged. Air-borne radar affords manifold potential uses in terrain clearance, collision prevention, and increased flexibility of navigation.

AIR-TRANSPORT-COMPANY FACILITIES

To supplement the government air navigation facilities, traffic control, and communication services, the air-transport companies have installed two-way communication equipment at approximately 100- to 200-mile intervals along the airways of the nation and in all their airplanes. This system permits continuous contact between the offices of each company and their aircraft in flight, and is utilized for transmission of CAA flight instruction, auxiliary weather and advisory information, and normal details of transport operation. The transport companies also operate teletype circuits and point-to-point radio stations which provide the rapid communication between operating offices which is essential to the successful operation of high-speed passenger, mail, and express service. The facilities of the different air-transport companies are coordinated through Aeronautical Radio, Inc., an association organized for this purpose and having its headquarters in Washington, D.C. The radio facilities already enumerated are sufficient for air-transport companies operating over the civil airways of the United States. In the case of international routes, corresponding facilities are required and these are also coordinated through Aeronautical Radio, Inc. Foreign representation in this association include airlines operating in Europe, South America, and Australia.

First organized for effecting joint action in the procurement of radio licenses from the FCC, Aeronautical Radio, Inc., now coordinates not only communications requirements but also the requirements of air and ground electronic equipment and assists in disseminating specifications information to various manufacturers in order that equipment produced will meet the characteristics required by the aviation industry. It makes available to the airlines weather data obtained from the Weather Bureau and

from its own local stations; and in cooperation with the airlines keeps track of aircraft, interprets weather data, etc. It is also capable of conducting the operation of all ground communications equipment at an airport serving a particular city, as in the case of Boston, where it now operates all such facilities. Its research facility, the Airborne Instruments Laboratory, conducts centralized research on problems of particular interest to the air-transport industry, *e.g.*, traffic control, collision prevention, antennas.

The Air Navigation and Traffic Control Committee of the Air Transport Association (ANTC), organized in 1945, represents the air-transport industry in the evaluation of air-navigation and traffic-control systems and in the planning of long-range projects. Its findings are submitted to the Airborne Instruments Laboratory for the necessary research.

The Radio Technical Commission for Aeronautics (RTCA) is an important organization serving the entire aviation industry and represents CAA and other air-line agencies of the government, the Armed Forces, the air lines, the aircraft radio manufacturers, etc. This Commission, formed in 1935, operated as an informal organization until 1942, at which time it adopted its original constitution. It was subsequently formally reorganized in January, 1947. Its present objective is to advance the art and science of aeronautics through the investigations of all available or potential applications of the art of telecommunication, to coordinate these applications with those of allied arts, and to adapt these applications to operational requirements. The Commission conducts studies to determine the suitability of existing and proposed systems of aids to navigation, communication, and traffic control. It also fosters new developments to meet aeronautical operating requirements, serves as a means of coordinating the views of government and industry on matters within its jurisdiction, and formulates recommendations based on its findings.

14. Radio Frequencies in Civil Aviation. The frequency requirements of aviation radio communication facilities are similar to the requirements of corresponding facilities in other applications and will not be specially considered here. The frequency requirements for radio navigational aids extend through the presently usable portion of the spectrum and are exemplified by the data in Table 2. These data are based upon a comprehensive outline of requirements prepared for use of member states by the Aeronautical Telecommunication and Radio Aids to Air Navigation Division (GOM) of ICAO. On July 1, 1947, the International Telecommunications Union convened in Atlantic City, N.J., to consider the reallocation on an international basis of frequency bands to all types of radio service, including radio aids in aviation.

The choice of operating frequency for a particular navigational aid depends upon a number of factors: wave-propagation characteristics, atmospheric noise level, bandwidth requirements, directional pattern requirements, and need for frequency adjacency to other aids so as to simplify the air-borne equipment problem. An understanding of propagation characteristics is essential to an understanding of the operational properties of many of the systems. Expansion of this discussion to include the influence of the other factors enumerated upon the present distribution of radio frequencies for navigational aids is impracticable here. However, a few words on the more general considerations will be attempted.

In the early organization of facilities on the civil airways of the United States, the frequency range, 200 to 400 kc, was chosen for the directional radio ranges, auxiliary voice and markers facilities, and the frequency range, 3,000 to 7,000 kc, for air-line two-way communication. Shortly before the war, there was a distinct trend to the adoption of v.h.f. for all services except direction finding, and v.h.f. was being applied partially for the latter service. The reasons directing the shift were substantially as follows:

Atmospheric disturbances arising from electrical storms constituted a severe limitation to reception in the 200- to 400-kc band and to somewhat lesser degree in the 3,000- to 7,000-kc band. Another form of disturbance, called *precipitation static* and of importance only in aircraft reception, constituted a second, and often even more serious, limitation to

Table 2. Proposed R-f Bands for Radio Aids to Air Navigation*

Service	Frequency	System	Present status of frequency requirement
1. Long-range facilities...	10-20 kc 70-100 kc 180 kc 300 kc 255-415 kc	V-l-f Decca (British) L-f omnirange L-f loran P.O.P.I. (British) ADF-consol (Europe)	Developmental Developmental International Developmental Regional
2. Short-distance aids and ranges	1,800-2,000 kc 200-415 kc 112-118 Mc 200-300 Mc 200-415 kc 255-415 kc 173-177 Mc	Standard loran Existing ranges (U.S., etc.) V-h-f omniranges Multitrack range (Australian) Markers, etc. (U.S., etc.) European navigational aids Homing on surface vessels (including DME)	Regional Regional International Regional Regional Regional (temporary)
Locater beacons and ground d.f.	200-235 Mc 200-300 Mc 1,000 Mc	Miscellaneous DME (ICAO std.) DME (U.S.)	Regional (temporary) International Regional
A.S.V. beacons.....	70-130 kc	L-f Decca (British)	Regional and development
Distance-measuring equipment	20-40 Mc 60-85 Mc 420-460 Mc	Gee (British)	Regional
Position-finding systems		Pulse-phase azimuth system (British)	Developmental
Radar surveillance....	1,600-1,700 Mc 2,700-3,300 Mc 2,700-3,300 Mc 8,500-10,000 Mc	Miscellaneous GCA search function Misc. (for air-traffic control) Air-borne (for air-traffic control)	Developmental International Developmental Developmental
Air-borne responders..	2,700-3,300 Mc 8,500-10,000 Mc	* †	
3. Final approach and landing facilities			
Localizers.....	108-112 Mc	ILS localizers	International
Marker beacons.....	200-415 kc	M-f locators	Regional
Approach beacons and glide paths	74.6-75.4 Mc 33.0-44.0 Mc 328.6-335.4 Mc 5,000-5,250 Mc 8,500-10,000 Mc	ILS marker beacons Standard beam approach ILS glide path Microwave system (ILS) CMR precision control	International Regional (temporary) International Developmental International
4. Aeronautical mobile and d-f services (including airport utilities)...	118-132 Mc	Miscellaneous	International
5. Distress facilities.....	500 kc	M-f direction finding	International
	†	H-f direction finding	International
6. Altimeters§.....	420-460 Mc 1,600-1,700 Mc 4,200-4,400 Mc	Miscellaneous Miscellaneous Miscellaneous	Regional Developmental Developmental
7. Integrated systems....	Several comprehensive systems, such as Teleran and Navar, are in various stages of development and utilize one or more sets of frequencies in the following regions: 100 Mc, 1,000 Mc, 3,000 Mc, 5,000 Mc, and 10,000 Mc.		

* Based upon a comprehensive outline of frequency requirements prepared for use of member states by the Aeronautical Telecommunication and Radio Aids to Air Navigation Division of the ICAO.

† Existing air-borne responders operate at 3,246 to 3,466 Mc and 9,300 to 9,320 Mc. ICAO has as yet made no recommendation as to standardization of frequencies. The use of responders in this band in conjunction with surveillance radars would necessitate international standardization of frequencies for interrogation and response.

‡ A frequency in the h-f band to be assigned by the International Telecommunication Union. The exact frequency will require agreement between aviation and maritime services.

§ ICAO has as yet made no recommendation as to the standardization of altimeter frequencies, but recognizes the temporary utility of the band, 420 to 460 Mc for this purpose, inasmuch as a limited-range altimeter already operates in this region. The other two bands are recognized as of potential utility for altimeter operation.

reception at these frequencies. This form of disturbance accompanies rain, snow, and even sandstorms and is caused by oscillating corona discharge from points on the airplane to the surrounding atmosphere. Its intensity is often sufficient to paralyze the ground-to-aircraft service at the frequencies considered. Accumulated experience indicated that reception on v.h.f. was practically free from atmospheric disturbances and was less influenced by precipitation static.

V.h.f. also afforded freedom from dependence upon ionospheric conditions. At the lower frequencies, range-course swinging is produced by the sky wave at night and is often the limiting factor on the useful distance range, particularly in mountainous terrain. Ionospheric propagation renders air-borne direction finding impracticable at the lower frequencies (at night) and at higher frequencies both day and night. Such direction finders employ the loop antenna because of its simplicity, but, unfortunately, this antenna is particularly susceptible to polarization errors from sky-wave transmission. V.h.f. further afforded the possibility of greater directivity of transmission and reception. At the lower frequencies, the directive patterns available for radio range utilization comprised the figure of eight and the cardioid. In forming a course with such patterns, considerable radiation existed in directions at large angles to the course with the possibility of reradiation from natural or artificial objects resulting in bent and split courses. The availability of patterns of greater directivity provided a means for reducing such effects. Since the ground wave from a v-h-f radio range is attenuated very rapidly, course bending due to diffraction of the ground wave in passing over different terrain is also eliminated.

Although the reasons cited have in part been counteracted by new system developments and by new operational requirements on the airways, the shift to v.h.f. is evident from a summary of the frequency recommendations of the C.O.M. Division in the band 108 to 132 Mc (see Table 3). The services now afforded a private aircraft equipped

Table 3. Proposed Frequency Allocation Plan for the V-h-f, 108- to 132-Mc Band

Band frequency, Mc	No. of channels	International service
108.1-111.9*	20	Air navigation aids; instrument-landing localizer with simultaneous radiotelephone channel
112.1-117.9	30	Air navigation aids; v-h-f directional and omnidirectional ranges with simultaneous radiotelephone channel
118.1-121.3	17	Air-traffic-control communications; ground to air only (with circularly polarized transmission)
121.5†	1	Emergency
121.7, 121.9	2	Airport utility; simplex or cross band (includes taxi instructions to aircraft on ground)
122.1, 122.3	20	Private aircraft en route; air-to-ground
122.5-122.9	3	Private aircraft to towers; air-to-ground
123.1-123.5	3	Flight test and flying schools; (shared) simplex
123.7-125.5	10	Approach control; air-to-ground
125.7-126.5	5	Air carrier; aircraft to towers
126.7	1	Air carrier; aircraft to airways stations
126.9-131.9	26	Air carrier en route; simplex

* In the United States, the frequencies 111.1 through 111.9 Mc have a secondary use as airways channels when the simultaneous voice channel of the v-h-f range is not adequate to provide for the traffic.

† Protected on either side by a guard band of 150 kc.

with a v-h-f receiver, 108 to 122 Mc, and a v-h-f transmitter, 121.1 and 121.5 Mc, include the v-h-f range guidance, weather broadcasts and flight advisories, two-way communication with the airway control centers, and two-way communication, in the air or on the ground, with airport control tower. The use of the runway localizer unit of ILS is also afforded. Full use of ILS would require the addition of a marker and landing-beam receiver. However, the localizer could be used in conjunction with the GCA instructions received from the tower.

Despite the concentration of certain of the facilities in a narrow portion of the v-h-f spectrum, an important new factor in air-transport operations requires the continued

use of the lower frequencies. This is the need for communication and navigational facilities over longer distances. Although the distance range of the v-h-f facilities may be increased appreciably by operating the aircraft at higher altitudes, the distance ranges involved dictate the utilization of relatively low frequencies (based primarily on ground-wave propagation) for the navigational aid and high frequencies (depending primarily on ionospheric propagation) for the communication service.

The l-f omnidirectional range, standard loran over water and l-f loran over land, and the Consol (or Sonne) systems, all have as a common characteristic the reliance on the ground wave for their primary service area. Each system provides a varying degree of extension of the service area through ionospheric propagation. The effects on accuracy of the navigational service in these secondary areas will be treated in detail in connection with the technical descriptions of the individual systems.

Long-distance communication, using high frequencies, may be supplemented with a ground-direction finding service, since the ground direction finders may employ complex antenna systems to reduce polarization errors.

RADIO NAVIGATIONAL SYSTEMS

15. Functional Classification. The various navigational systems may be classified according to the shapes of the position lines they generate: *radial* for the radio ranges and for the d-f systems; *hyperbolic* for the time-difference methods; *circular* for the distance-measuring responder systems. Each class may be applied to short- or long-distance service, depending upon the choice of operating frequency; however, the choice will in turn introduce operational characteristics that may have important bearing upon the utility of the system. The classification as to the shapes of the position lines is more basic than that of type of service and is used in organizing the following technical material on navigational systems.

Unfortunately, the terminology developed for radio navigational systems has become somewhat anomalous so that the reader must often decide what is meant on the basis of his own experience. Thus certain types of radial systems are called "radio ranges," hyperbolic systems are termed "differential range systems," and circular systems are referred to as "pure range systems."

16. Radial systems depend upon the directional radiation patterns (in the horizontal plane) of simple antenna systems to delineate radial directions in space.

In the simplest form, the figure-of-eight pattern of a loop antenna with its plane vertical (or of two nondirectional radiators spaced $\lambda/2$, or less, apart and driven 180 deg out of phase) is employed. The phase and amplitude characteristics of this elementary radiating system are shown in Fig. 2, for a spacing of exactly $\lambda/2$. The solid lines represent loci of equal phase difference between the radiation from antenna *A* and that from antenna *B*. The specific phase differences are indicated in the individual lines. These lines are obviously hyperbolas with the antennas as foci. At distances from the center of the base line (connecting the two antennas) greater than five times the length of the base, the hyperbolas are essentially straight lines radiating out from the center of the base line. The hyperbola that is the bisector of the base is exactly a straight line at all distances. Since, at the frequencies used in aviation radio, $\lambda/2$ is never very great (1,640 ft for 300 kc), all the hyperbolas may be considered straight for practical distances.

Along the line marked 180 deg, there is exact cancellation of the fields from the two antennas, whereas maximum addition of the fields occurs at 0 and 360 deg. The horizontal radiation pattern from the two antennas is thus as shown by the dotted (figure-of-eight) curve of Fig. 2. The loci of constant phase difference and the horizontal radiation pattern for the same pair of antennas, spaced exactly $\lambda/2$ apart but in phase synchronism, are shown in Fig. 3.

An analysis of the utility of the two antenna arrangements for d-f and radio range purposes will serve to bring out certain important basic principles. Let ϕ be the phase angle, E the amplitude of radiation, and θ the azimuth angle. Then $\Delta\phi/\Delta\theta$ and $\Delta E/\Delta\theta$ will be seen to be highest in the region near phase cancellation for the differen-

tial antenna connection (Fig. 2). This condition will also exist for antenna spacings less than $\lambda/2$, the figure-of-eight pattern becoming more nearly a true cosine figure as the spacing approaches zero as a limit. Null-type direction finders using a rotatable directional antenna (either a loop or an Adcock) take advantage of these properties of the differential antenna connection, *i.e.*, both the sharp null for determining the direction of arrival of the received wave and the small spacing for securing antennas of practicable dimensions, particularly for the lower frequencies. Air-borne automatic direction finders superpose the directional pattern of the differential antenna upon a nondirectional pattern, successively adding and subtracting the two and comparing

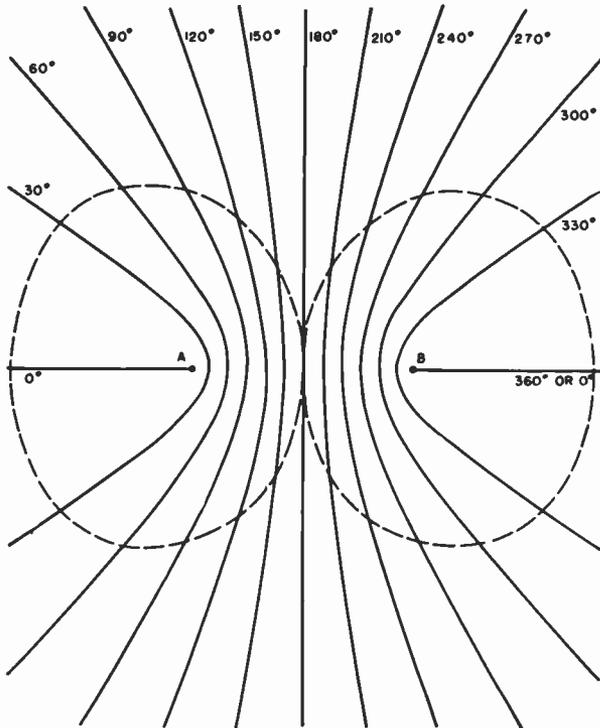


FIG. 2. Lines of constant phase difference and radiation pattern in horizontal plane for two radiators spaced $\lambda/2$ apart and in phase opposition.

the difference of the two results, thereby producing a "course" along the null of the differential antenna.

Fixed ground-station direction finders using cathode-ray indicators also require the differential connection of Fig. 2. Two pairs of antennas are crossed at right angles with each pair feeding the stator winding of a goniometer. The two stator windings are also crossed at right angles, and one or more rotors (search coils) connect to the d-f receiver. The coupling of each rotor to each of the stators is sinusoidal so that each rotor becomes the equivalent of a rotatable loop antenna. This can occur only if each antenna pair yields a true cosine pattern which is a characteristic of the differential antenna connection with small spacing.

For radio range-beacon use, two figure-of-eight patterns are frequently crossed to produce "equisignal" courses along the radial lines from the center of antenna system through the points of intersection of the two patterns. At relatively low frequencies,

it is convenient to use the differential connection in order to avoid large antenna spacings. Even if it were convenient to use the synchronous connection of Fig. 3, the intersection of the two patterns would occur at lower pattern amplitudes and result in effective power loss. It will be apparent that, when two patterns are overlapped to produce an equisignal course, the important parameters are the ratio of the equisignal field to the maximum field and the rate of change of field strength with azimuth angle in the region near the equisignal.

Certain radio-range beacons, including the omnirange, employ a rotating figure-of-eight pattern which can be produced more accurately with small antenna spacings.

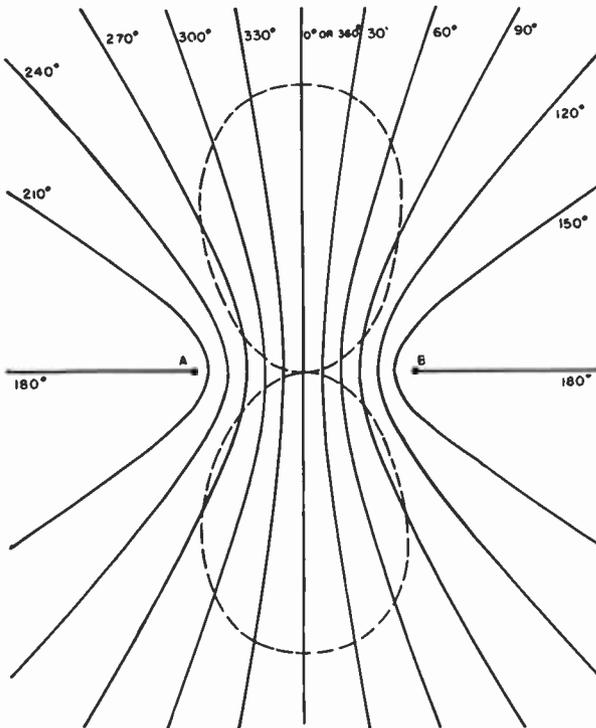


FIG. 3. Lines of constant phase difference and radiation pattern in horizontal plane for two radiators spaced $\lambda/2$ apart and in phase synchronism.

This is the exact reverse of the fixed d-f case just described. The pattern may be rotated at any desired rate including the carrier frequency rate, although the latter rate has so far had limited utility.

For the synchronous antenna connection (Fig. 3) the distribution of hyperbolas is identical with Fig. 2, but the corresponding phase-difference values are modified because of the different phase relationship between the two antennas. Thus, $\Delta\phi/\Delta\theta$ and $\Delta E/\Delta\theta$ are highest near the region of phase addition, so that the figure-of-eight pattern has a sharper maximum but much broader minimum. If the spacing between the two antennas is increased by a factor of 2, $\Delta\phi/\Delta\theta$ also increases by 2 so that the number of regions of phase addition and opposition is also doubled (see Fig. 4). The radiation pattern now has four lobes, as shown.

Great flexibility is available for securing a variety of radiation patterns by varying the antenna spacing; additional degrees of freedom are afforded by varying the

relative amplitude and phase of the currents in the two antennas. This principle is utilized, for example, in producing two multilobed patterns for overlapping to provide 24 equisignal courses, as shown in Fig. 5. The antenna spacing is 3λ , the ratio of currents in the two antennas is 2:1, and the phase difference is 90 deg. The solid-line pattern is shifted to the dotted-line pattern by reversing the phase of the current in one antenna. It will be evident that $dE/d\theta$ (and hence, course sharpness) is proportional to $N \cos \theta$, where N is the antenna spacing in wavelengths, and θ is measured from the perpendicular bisector of the base line containing the two antennas.

17. **Hyperbolic Systems.** Such systems depend upon the phase difference or time difference of arrival of waves from two antennas (disposed along a base line) to delin-

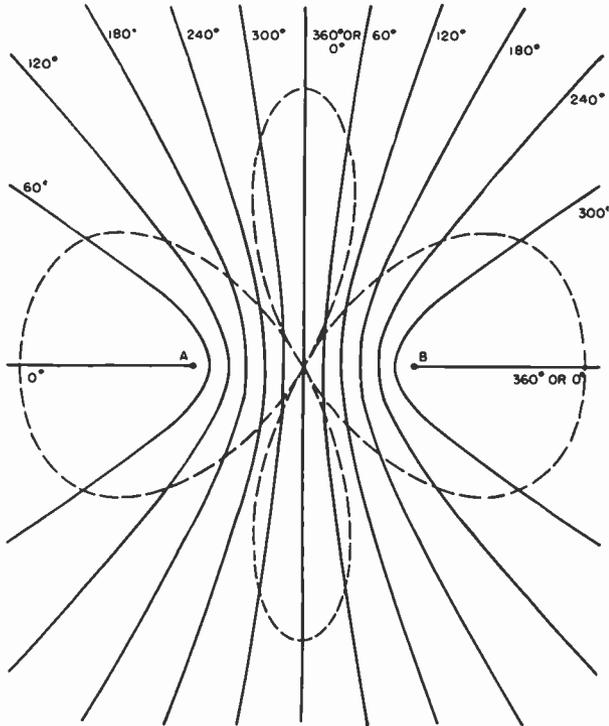


FIG. 4. Lines of constant phase difference and radiation pattern in horizontal plane for two radiators spaced λ apart and in phase synchronism.

eat lines of position in space. Figures 3 and 4 depict the situation for phase-difference measurement for a base line of $\lambda/2$ and λ , respectively. Measurement of the phase difference of two continuous waves is practically impossible; thus a variety of expedients are adopted. One is to interrupt the transmissions from the antennas successively, to transmit a nondirectional reference carrier of slightly different frequency continuously, and to compare (on a memory basis) the phases of the two audio signals produced by the beating in the receiver of the reference carrier with the successive transmissions from the two antennas. It is well known that the phases of the two transmissions will be preserved in the phases of the corresponding audio signals if the reference carrier is phase-controlled.

A second expedient is to transmit on each of the two antennas a different carrier frequency, the two frequencies being so related that their multiplication by different

small integral numbers, say 3 and 4, will produce the same frequency. At the receiving end, the received transmissions are frequency-multiplied to a common frequency and their phases compared in a discriminator circuit.

A third expedient is to transmit on a different frequency from each of the two antennas but to modulate both transmissions with a single audio frequency such that the modulation envelopes of the two transmissions leave the two antennas in phase synchronism. If the base line is chosen as 180 km and the modulation frequency 833 cycles, the base line is then equal to half the period of the modulation frequency or

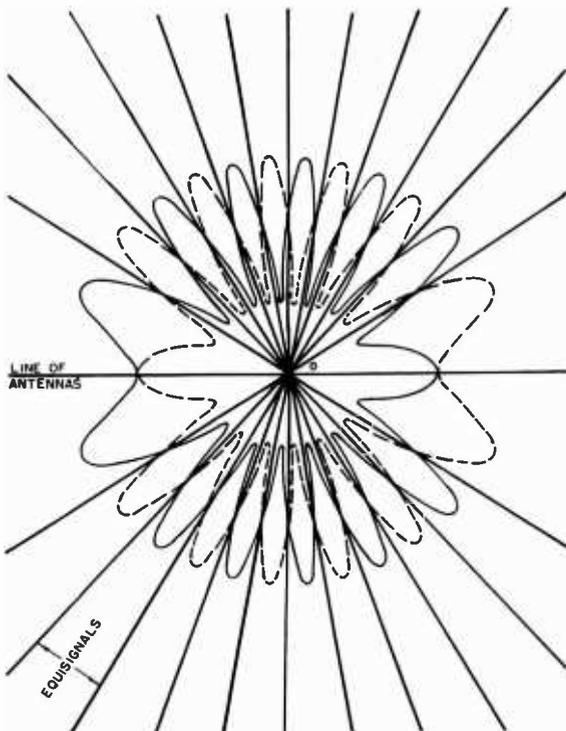


FIG. 5. Multiple-course equisignal pattern formed by two radiators spaced 3λ apart with current ratio of 2:1 and phase difference of 90 deg. (Phase in one antenna alternately reversed to produce alternate patterns.)

$\lambda/2$. $\left(\lambda = \frac{3 \times 10^5 \text{ km/per sec}}{833 \text{ cps}} = 360 \text{ km.} \right)$ Figure 3 then applies, and it will be seen that the spacing between adjacent hyperbolas represents a difference in distance of travel of the two waves of 30 km, corresponding to a phase difference of 30 deg. If the modulation frequency is increased to 1,666 cycles and the base line kept as 180 km, the antenna spacing for the modulation frequency is a full wavelength, λ (see Fig. 4). Since $\Delta\phi/\Delta\theta$ has now doubled, it will be evident that a given hyperbolic line of position may be determined (by phase measurement) with twice the precision.

For the lower modulating frequency (833 cycles), the time of travel corresponding to 1 wavelength of the modulation envelope is $10^5/833 = 1,200 \mu\text{sec}$. Since pulse techniques permit the accurate forming of pulses so that time differences to an accuracy of the order of 1 μsec may be measured, whereas it is difficult to measure phase difference to better than about 1 deg, there is a superiority, for the case assumed, of

better than 3 to 1. Increase of the modulating frequency or phase measurement at the carrier frequency can make up for this difference in measurement and, in fact, provides potential accuracies much better than for the pulse methods. However, the pulse method affords great simplification in that it becomes possible to distinguish the radiation from the two antennas on a simple basis, *i.e.*, time of arrival.

Figure 6 shows the situation for a 600- μ sec spacing of the antennas, *i.e.*, 180 km. Here a pulse emitted from station *B*, the master, arrives at station *A*, the slave, in 600 μ sec and sets off a second pulse after 700 μ sec of delay. The pulses from *B* will thus arrive ahead of the pulses from *A* at all points in space, so that each hyperbola is

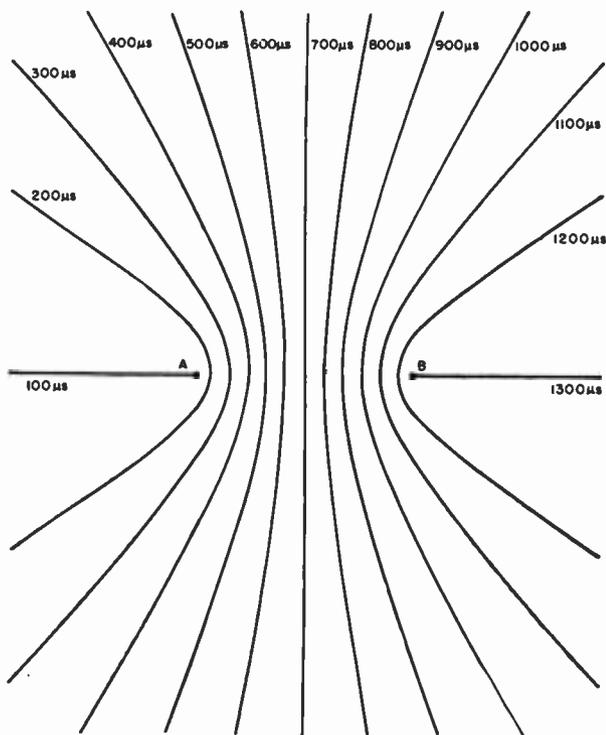


FIG. 6. Family of hyperbolic lines of position formed by two antennas 600 μ sec apart with the *A* pulse delayed 700 μ sec behind the *B* pulse.

defined unambiguously by measuring the difference in time of arrival between the first and second pulses.

It is well to note at this point that, for the carrier phase-measuring systems, the base line can be quite short in terms of distance so that the hyperbolas become radial lines at virtually all practical distances. On the other hand, the base line for the modulation-envelope phase-measuring systems and for the pulse-time systems needs to be long, so that the position lines generated are hyperbolas in fact. In the pulse-time systems, it is desirable to have very long base lines since the lines of position are then more nearly parallel over a large service area. Long base lines, however, introduce requirements for supporting transmission (generally over land) between the master and slave station and for introducing in the receiver the capability of measuring a wider range of time differences.

In the hyperbolic systems (often termed *differential range* systems) it is generally necessary to provide a second antenna arrangement to produce a set of hyperbolic lines of position substantially perpendicular to the first set. By determining his position on both sets of hyperbolas, the navigator determines a fix in space. There are notable exceptions, however, since the hyperbolas of one set of antennas, with rather short base lines, may be utilized as multiple radial tracks to the center of the base line.

Since time measurements and propagation path characteristics are subject to error, a particular hyperbolic line of position established by a pair of antennas has a degree of width in space. The intersection of two such lines produces a zone of uncertainty. The shape and extent of such a zone will depend upon the angle at which the two lines of finite width intersect. The accuracy of fix from two sets of hyperbolas is thus best in areas where the two position lines intersect at right angles and worst where the two position lines approach parallelism. The case for long base lines is again indicated.

18. Circular systems depend upon the measurement of distance to two or more geographically separated, fixed locations to determine a position fix. The navigator may radiate a signal to man-made or natural reflectors, or to radar beacons, and determine the distance to such reflectors by measuring the time taken by the signal to make the round trip to the reflector and return. He may, of course, also measure the direction of arrival of the returned wave, but this phase of the measurement becomes a radial system. The fix consists in drawing circles with the fixed geographical locations as centers and with radii equal to their measured distance from the navigator. Any two such circles will intersect at two points, one of which is the true position. The ambiguity between the two points may be resolved in a variety of ways, one of which is by distance measurement to a third, strategically situated, location.

An alternate procedure is for two or more geographically separated ground stations to make distance measurements to the aircraft by reflection or, preferably, by reemission from a radar beacon on the aircraft which also furnishes identification of the craft.

The circular system is seen to depend upon radar techniques. There are a number of equipments for measuring distance, termed DME (distance-measuring equipment). Most of these use pulse-time measurements, but a few employ phase differences in a modulation wave envelope.

RADIAL NAVIGATION SYSTEMS

19. CAA L-f Directional Radio-range Stations with Simultaneous Voice Transmission (RA: 200-415 Kc). Developed a decade before the war, this system still bears the brunt of furnishing directional guidance along the civil airways of the United States. The transmitter employs two independent r-f channels controlled by two matched A-cut quartz plates 1,020 cycles apart. Continuity of service is ensured by means of complete stand-by equipment provided with an automatic transfer relay. The antenna system comprises five self-supporting base-insulated steel towers approximately 130 ft high, four of which are placed at the corners of a square 300 to 500 ft on a side, and the fifth at the geometrical center of the square. These antennas, which are now used at 235 range stations, take the place of the older loop antennas which are still used at 125 stations. One of the r-f channels of the transmitter delivers 400 watts of carrier power (which may be modulated 70 per cent by speech) to the center antenna. The other delivers 275 watts of unmodulated carrier power to a coupling system which feeds the four corner antennas. In the absence of speech modulation, the setup forms a single-side-band system having 1,020-cycle modulation. The carrier is radiated nondirectionally by the center tower; whereas, the side band has the characteristic radiation of the radio range beacon (see Fig. 7a). When special modulation is applied, the central tower radiates, in addition, the speech side bands, which are also nondirectional. The system affords means for the simultaneous radiation of weather broadcasts and directional guidance signals. To avoid interference between the 1,020-cycle beacon signals and the speech frequencies, a band-rejection filter is inserted in the input circuit to the speech modulation for eliminating the speech frequencies in the region of 1,020 cycles. A combination band-pass band-rejection filter

is used in the output circuit of the aircraft receiver, so that one circuit carries only the range beacon (1,020-cycle) signals, and the other circuit carries the speech signals.

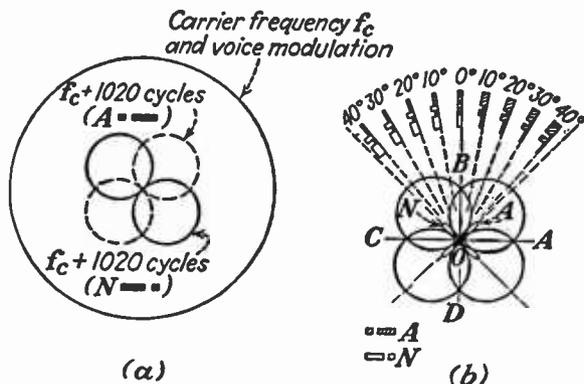


FIG. 7. (a) Space radiation pattern of simultaneous 1-f directional range-beacon and voice broadcast station. (b) Range-beacon signals produced.

in the sky wave, leading to random errors in the indicated beacon courses, is minimized.

20. Range-beacon Operation. The principles of operation of the radio range beacon, whereby radio marked courses are set up, are evident from Fig. 7b. The intensities of the side-band emissions formed by the two directional antenna systems, and hence of the intensities of the detected signals produced by beating the side-band emissions with the nondirectional carrier, are equal along the lines OA , OB , OC , and OD , which bisect the angles between the two antenna systems. An aircraft may therefore follow a course along the bisectors referred to, if means are provided for distinguishing the radiations from the two directional antennas. For this purpose, an automatic keying relay, connected in the coupling circuit from the side-band channel of the transmitter to the directional antenna system, is used for keying the radio power to one of the directional antennas in accordance with the Morse character N (— ·), and the power to the second directional antenna in accordance with the Morse character A (· —). The coded signals are sent out in groups and are interlocked, so that along any one of the four courses they form a long dash, or continuous monotone signal, interrupted every 24 sec by the station identification signal. The course signals are obtained along zones 2 to 3 deg wide. Off the course, the monotone signals break up into the component A and N signals, one or the other being of greater intensity, depending upon the side off course. The pilot is thus enabled to return to the course if the aircraft should drift to either side for any reason.

The coupling system between the transmitter and the directional antennas incorporates the link-circuit relay, a goniometer, a course-shifting pad, artificial-line sections, concentric transmission lines to the tower antennas, and antenna coupling and tuning equipment (see Fig. 8). The relay is energized by an automatic motor-driven keying device (not shown), so as to key the r-f power to the primary windings P_1 and P_2 of the goniometer, in accordance with the N-A sequence indicated in the foregoing description. The goniometer offers a convenient method of orienting the beacon space pattern, and the course-shifting pads and artificial-line sections are used for shifting the range-beacon courses from their 90-deg relationship, in order that they may be aligned with the airways. Precautions, in the form of special automatic phase and amplitude control of the antenna currents, are provided to ensure maintenance of the space patterns, so that drifting of the courses, due to changes in tower capacitance under varying weather conditions, will not exceed 1.5 deg.

Although the 1-f four-course radio range is a simple type of navigational aid, it has several marked disadvantages. The limited number of ground tracks available frequently necessitates excessively long circuitous flights in point-to-point travel. Several factors contribute to producing a poor signal-to-noise ratio in the 1-f bands, such

as (1) poor antennas, (2) precipitation static caused by rain, snow, and dust, (3) atmospheric noise, (4) high receiver noise, and (5) locally generated noise. Pilots experience fatigue in listening to the monotonous tones in the A, N, or twilight T zones. Swinging of the course occurs under night conditions and is not entirely eliminated, partic-

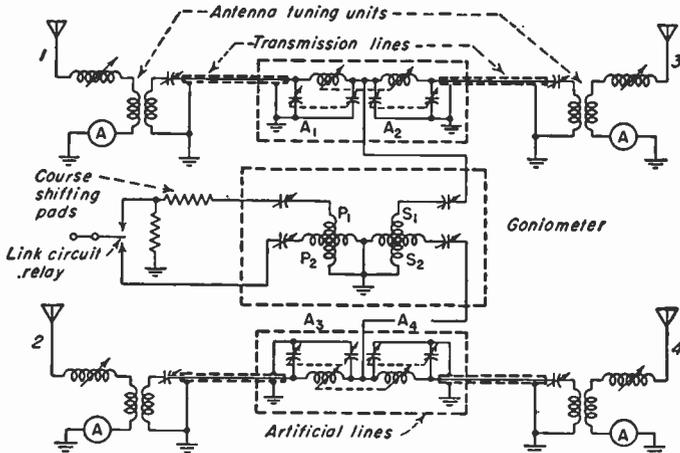


FIG. 8. Schematic diagram of range-beacon antenna coupling system.

ularly in mountainous terrain. Course ambiguities can be resolved only through complex orientation procedures which place too great a burden upon the pilot with respect to both calculations and the exercise of personal judgment.

21. V-h-f Two-course Range Beacon with Visual Indication and Aural Sector Identification (VAR). This is an interim navigational aid on the U.S. civil airways, the characteristic pattern of which is shown in Fig. 9.

Only two useful courses are provided by the intersecting space patterns of the v-h-f two-course range, as compared with four courses in the case of the l-f ranges. This

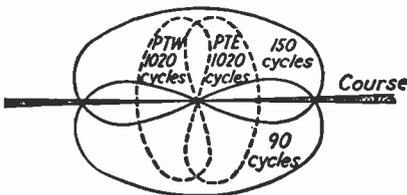


FIG. 9. V-h-f visual two-course range with aural sector identification.

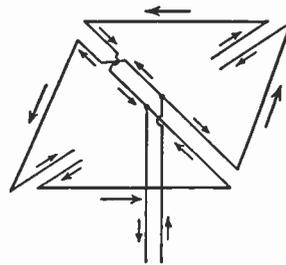


FIG. 10. V-h-f loop-antenna element for use in antenna arrays producing horizontally polarized waves.

simplification materially reduces the orientation problems which the pilot is frequently called upon to solve under special conditions. For example, with the four-course range, he may drift into an N or A quadrant, when near the station during strong winds, and experience difficulty in determining which N or A quadrant he is in and in which direction to fly to get on the desired one of the four courses.

The principle of this system was described in Sec. 4. The patterns shown in Fig. 9 are produced by five-element antenna arrays, employing horizontal antenna elements

of the Alford type. A single element, shown in Fig. 10, is equivalent to a horizontal loop antenna. Experience indicates that the use of horizontally polarized waves minimizes site errors and bent and multiple courses in mountainous terrain.

22. V-h-f Omnidirectional Radio Range. The operation of the omnirange can be likened to that of an airways light beacon revolving at a uniform rate. If the beacon makes 1 rpm and if a fixed red light on the beacon tower is caused to flash as the beam points north, an observer with a stop watch can determine his bearing merely by starting the watch when the red light flashes and stopping it when the beam flashes past him. The azimuth angle measured from north is then directly proportional to the elapsed time recorded by the stop watch.

The omnirange functions through the radiation of two separate r-f field patterns. One of these, a variable-phase pattern, corresponds to the rotating beam of the beacon. The other, providing a signal of reference phase, corresponds to the flashing red light on the beacon tower. These field patterns are produced by an antenna array consisting of four horizontal loop antenna elements (of the Alford type) located at the corners of a square, and a fifth element located at the center. The center element is excited with v-h-f energy (112 to 118 Mc), which is amplitude-modulated 30 per cent with a 10-kc¹ signal. The 10-kc signal is frequency-modulated with a 30-cycle reference signal to ± 480 cycles above and below 10 kc. In addition, voice modulations not exceeding 40 per cent can be applied to this same v-h-f carrier. By these means, the center element is caused to produce a nondirectional reference field pattern which also serves as a voice channel. The diagonally disposed pair of elements at the corners of the square are connected to a common feed point, but with different lengths of transmission line, so that the v-h-f energy applied to one element is 180 deg out of phase with that supplied to the other. The radiated pattern for this pair is thus approximately a figure of eight. The same arrangement is employed with the other pair of antenna elements diagonally disposed on the other two corners of the square. The transmitter supplies energy to these two pairs of antenna elements through a special, capacitor-type goniometer, the rotor of which is driven at 1,800 rpm in phase synchronism with the reference generator. The goniometer functions like two mechanical side-band generators, suppressing the carrier, and feeding two 30-cycle modulated side bands, displaced by one-quarter period of the modulation envelope, to the two pairs of antenna elements.

Since the modulation envelopes of the individual radiation patterns produced by the two pairs of side-band antenna elements are in phase quadrature, their resultant produces the equivalent of a rotating field. The field pattern for the north-south pair may be expressed as

$$e_1 = k \sin (S \cos \theta)$$

and for the east-west pair as

$$e_2 = k \sin (S \sin \theta)$$

where S is half the electrical spacing between the antenna elements and θ is the azimuth bearing from north. The resultant side-band field at any point P is then

$$E = \sqrt{2} K \tan^{-1} \frac{\sin (S \sin \theta)}{\sin (S \cos \theta)}$$

and if S is small, the phase of the side-band field is equal to the azimuth bearing as measured from north. With the actual antenna spacing employed, 56.1 deg, an octantal variation in the phase of about 2.5 deg amplitude (about a true sine-wave distribution) exists. This error may double if the two side-band voltages differ in amplitude by as much as 10 per cent. The azimuth angle of point P is derived in the receiver output by comparing the phase of the demodulated side-band signal with the phase of the demodulated carrier reference signal.

¹ Actually 9,960 cycles.

Omnirange Transmitter. A block diagram of the omnirange transmitter is shown in Fig. 11. Both the carrier modulation and the side-band modulation are carried out at the final power-output level. The carrier modulator is a $\lambda/4$ transmission-line type. The principle of the rotating capacitor-type goniometer is shown in Fig. 12. Two rotor plates, insulated from each other, form half cylinders with the ends tapered off. When rolled out flat, the shape of each plate is a double half-sine wave. Two sets of stator plates, each stator consisting of two insulated half-cylindrical shells, are

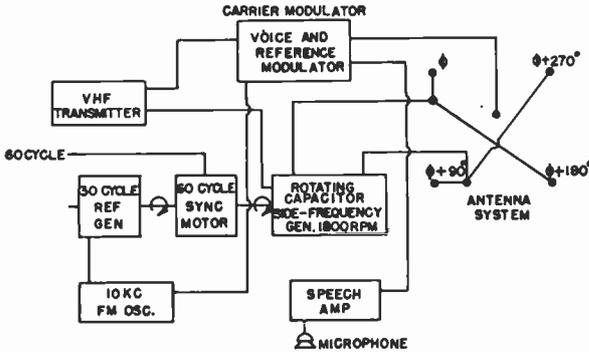


FIG. 11. Schematic of transmitter for v-h-f omnirange.

disposed side by side about the rotor, so mounted that the slits between the two plates of one stator are displaced by 90 deg around the circumference from the slits between the two plates of the second stator. V-h-f voltage from the transmitter is applied across the two rotor plates, and each stator feeds one pair of differentially connected antenna elements. The capacitor goniometer thus produces two v-h-f output amplitudes proportional to the sine and cosine of the displacement angle of the rotor.

V-h-f Air-borne Receivers. A block diagram of the special azimuth converter required in the output of the aircraft receiver for securing course indication along any

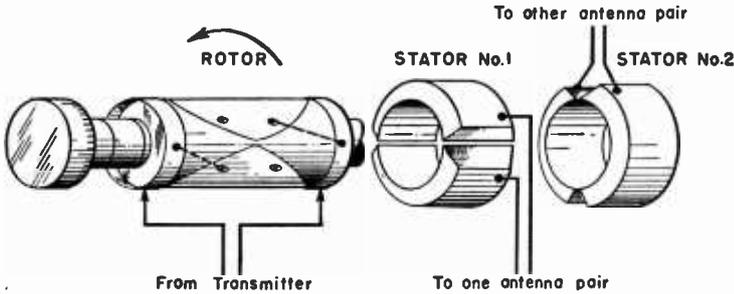


FIG. 12. Rotating capacitor-type goniometer for v-h-f omnirange.

radial of the omnirange is shown in Fig. 13. The function of this converter unit is to take the complex a-f signal voltage from the audio amplifier of the receiver and to recover the reference and variable-phase 30-cycle components so that they may be compared to yield the bearing of the receiving point with respect to the omnirange station and to provide, in addition, pointer-type course indications suitable for flying a course along the radial line corresponding to that bearing.

A-f voltage is taken from the cathode of the first amplifier in the receiver and applied to a complex filter circuit. The carrier of the reference-phase field beating with the side bands of the variable-phase field produces a 30-cycle voltage of variable phase

which is accepted by the 30-cycle low-pass filter¹ and fed to the channel termed "variable." The carrier of the reference-phase field beating with its own side bands produces a 10-kc subcarrier, frequency-modulated at 30 cycles, which is selected by the high-pass 10-kc filter and fed to the channel termed "reference." A discriminator in the reference channel recovers the 30-cycle reference-phase voltage from the 10-kc subcarrier and this voltage, amplified, is applied to one terminal of the meter-rectifier circuit for the vertical (zero-center) element of a crossed-pointer instrument.

In the variable channel, the 30-cycle variable-phase voltage is applied to a phase-splitting circuit which provides two equal voltages having a phase difference of 90 deg.

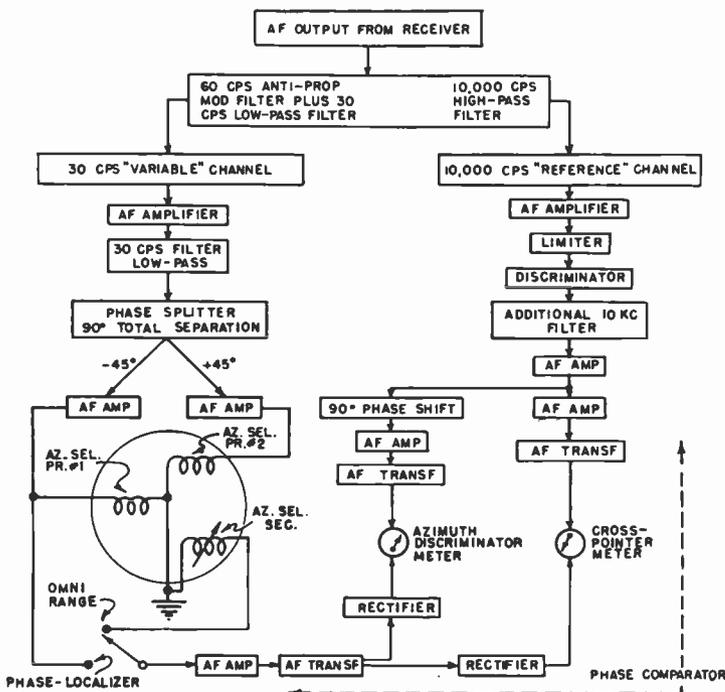


Fig. 13. Phase-comparison azimuth converter in v-h-f air-borne receiver. (Courtesy of Aircraft Radio Corporation.)

These two voltages are applied to the two stators of a goniometer-type phase-shifting device, designated "azimuth selector." The phase of the output voltage of this azimuth selector may be varied at will throughout 360 deg by varying the angular position of the goniometer rotor coil, with corresponding settings on a dial carried by the rotor calibrated in degrees of azimuth. The azimuth selector provides a means of obtaining on-course indications along any chosen radial course, merely by making the proper dial setting. The output of the azimuth selector is amplified and applied to the other terminal of the meter-rectifier circuit for the vertical (zero-center) element.

The two transformers, the rectifier, and the meter element form what is essentially a "wattmeter" circuit, so that the vertical pointer will read zero current, *i.e.*, "on course," whenever the two voltages applied from the two channels are 90 deg out of

¹ The 60-cycle antiprojection rejection filter is used to reject modulation of the received wave experienced occasionally when the airplane propellers approach critical positions between the receiving antenna and the omnirange station.

phase. Left-right deflection of the vertical pointer, if the proper setting of the azimuth selector has been obtained, will indicate that the airplane is to the left or right of the radial course to or from the omnirange. There is, however, the possibility of an 180-deg control ambiguity, since two 90-deg phase conditions exist. There is also the possibility of having the pointer read "on course" because of complete loss of signal.

Ambiguity Resolution. Indications to resolve ambiguity and to show loss of signal may be provided in several ways. A 360-deg azimuth indicator may be provided to show the aircraft's bearing from the station. This unit operates independently from the pointer instrument and is essentially a d-c selsyn-type motor (with pointer indicator) consisting of a permanent magnet rotor and two field coils at right angles to each other. A flag alarm on the combined instrument of the later types receives d.c. from the output of the reference channel to indicate presence or loss of signal.

An alternate method to resolve ambiguity and to indicate signal loss is shown in Fig. 13. A neon lamp or a miniature zero-center meter receives a voltage derived from a separate wattmeter which operates at a phase relationship of 90 deg from the wattmeter circuit that provides the course indications. When the neon lamp arrangement is used and the goniometer phase shifter is adjusted to give a zero indication on the left-right pointer of the crossed-pointer instrument, the voltages applied to the mixer amplifier will be either in phase or 180 deg out of phase. The correct setting of the phase shifter will produce the in-phase condition, thus lighting the neon lamp. Positive indication of signal presence for the on-course position is thus also obtained.

When the zero-center meter is employed for resolving ambiguity, a color scheme is used to provide the necessary correlation between the meter indications and the proper end of the pointer of the azimuth selector. The meter scale is colored green to one side of center, white to the other side of center, and red at the center proper; and the opposite ends of the azimuth pointer are colored green and white, respectively, to correspond with the coloring of the meter scale. The red area of the meter scale indicates the absence of either the reference signal or the variable-phase signal, or both.

The characteristics of a commercial air-borne receiver incorporating the azimuth converter are of interest. The v-h-f receiver is a tunable unit for reception of a-m signals in the band of 108 to 135 Mc. It employs a nine-tube superheterodyne circuit using a 4-section gang condenser and eight tuned circuits at 15 Mc in the i-f section. The r-f oscillator is below the dial frequency. A noise limiter and two a-f stages follow the detector. A.f. to the converter and to a filter-amplifier unit (for use with the two-course visual-aural v-h-f directional range) is taken from the cathode of the first amplifier stage. The latter unit is self-contained and may be omitted without affecting the normal operation of the remainder of the receiver. The amplifier is designed to cut off signals at the low and high ends of the a-f spectrum so that the voltage applied to the headset load in the amplifier output is down 25 db from maximum at 30 and 10,000 cycles. Speech, derived from beating of the carrier of the reference wave with its 40 per cent speech side bands, will thus be free of the reference and variable-phase omnirange signals.

In addition to providing the indications required for short range navigation (VAR and VOR), the same receiver is also capable of receiving transmissions from runway ILS localizers of either the equisignal or the phase-comparison types. Moreover, the aircraft indicator contains the necessary meters for following runway localizers and landing glide paths. The phase-comparison localizer transmits a nondirectional reference-phase pattern together with a fixed directional side-band pattern whose 30-cycle signal is in phase with the reference 30-cycle signal on one side of the runway and 180 deg out of phase on the other side. Exactly on the center of the runway, there is only the reference signal. By connecting one side of the phase-comparison wattmeter (see Fig. 13) to one primary of the azimuth selector, the omnirange converter may be used for securing course indications from the phase localizer. In practice the variable phase of the omnirange is adjusted to match the reference phase at 45 deg from north, so that no phase shift is experienced when going over to localizer reception.

The a.v.c. is the delayed type which allows the output to build up to approximately 170 mw before taking hold. The noise-limiter employs a series-diode circuit with provision for clipping signals above 80 per cent modulation. The a-v-c diode connects to the seventh i-f tuned circuit, leaving the last i-f tuned circuit to provide aural tuning in lieu of a visual tuning meter in certain applications of this receiver. Power output from the knee of the

a. v. c. at approximately $6 \mu\text{v}$, up to 100,000 μv input, rises from 170 to 360 mw, for signals modulated 30 per cent at 400 cps. Normal output load is 300 ohms. The sensitivity of the receiver is not poorer than $2 \mu\text{v}$ at any frequency in its range ($m = 0.3$, 400 cps, 10 mw into 300-ohm load). The selectivity (band width) is 100 kc for 6 db and 350 kc for 60 db. Particular attention has been paid in the receiver design to keep spurious responses to a minimum. The spurious response voltage ratios are not poorer than 1,000,000:1 for i. f., 40,000:1 for image, 30,000:1 for ($2 \times$ dial - 3 i. f.), 100,000:1 for (dial - $\frac{1}{2}$ i. f.).

The air-borne receiving antenna, used for the reception of omnirange signals, is a horizontal V type with an 80-deg apex angle and is fed by a coaxial line. Its most suitable mounting position is atop the vertical stabilizer where the altitude effect and the effect of propeller modulation have been found to be a minimum.

The omnirange system produces indications which are accurate to about ± 1.5 deg on radial courses free from serious reflections. Transmitting stations having a rating of 200 watts and properly located will provide coverage to about 100 miles for aircraft at low altitudes and up to about 250 miles for aircraft at altitudes of the order of 20,000 ft.

The site at which omnidirectional stations are located must be chosen to give adequate optical coverage, to be nearly free from hills higher than the site, and to be free from electrical irregularities, such as trees and gullies, within about 0.2 mile of the installation. The use of horizontal polarization results in material freedom from all but major site irregularities.

The five-loop antenna array is mounted on a rotatable base plate on top of a standard range tower (up to 50 ft high) surmounted by a circular counterpoise 35 ft in diameter. Surrounding the loop antenna array is a vertical polarization screen consisting of a cylindrical cage of vertical wires. Elimination of vertically polarized radiation eliminates "course pushing" occurring during turns of the receiving aircraft and attendant tilting of the horizontal receiving antenna.

The CAA does not permit the use of equipment for navigational aids in blind-flying conditions unless this equipment is monitored and some indication as to accuracy of the information is presented to the pilot. The monitoring system consists of a remote receiver and a phase-comparison unit. The remote receiver, housed in a weather-proof case, is located 200 ft from the station and 45 deg from the position where the reference and variable channels are in phase. The phase-comparison unit is located in the omnidirectional station and receives signal from the remote receiver over a single-conductor, coaxial, weatherproof cable. The phase-comparison unit is virtually identical to the converter unit of a v-h-f omnidirectional receiver. A meter is used to measure the output of both channels and a second meter to measure discriminator output or to serve as an indicator in the phase-comparison circuit. A calibrating circuit is incorporated comprising a mixer stage having one input from the 10-kc f-m oscillator and the other input from the 60-cycle generator. The output of this mixer may be compared against the output of the remote receiver. Once the monitor is in operation, any deviation in the phase relationship of the radiated reference and variable channels will actuate a sensitive relay in the phase-comparison circuit and may be used to give an alarm.

R/ θ Computers. Navigational computers are being developed to extend the utility of the omnirange beacon (when employed with DME) so that a navigator may secure automatic visual indication of distance and azimuth from any arbitrary point (within the coverage of the over-all omnirange system) and follow any operationally desirable track.

23. L-f Omnidirectional Radio Range. An l-f omnidirectional radio range now under development by the CAA is expected to provide a useful range of 600 miles or greater over land with an accuracy to within ± 3 deg. Although this system operates in the band 70 to 400 kc, it employs the same general phase-comparison principle as the v-h-f omnidirectional range. In the latest version, undergoing service tests at Nantucket, Mass., at the time of this writing, the same method of separating the variable- and reference-phase signals is employed as for the v-h-f omnirange. The var-

iable-phase field is produced by a carrier fed to the rotor coil of an inductor-type goniometer, the two stators of which (disposed at right angles) feed two pairs of diagonally disposed vertical towers. The rotor is driven mechanically at 1,800 rpm. Each stator thus drives its pair of towers to produce a figure-of-eight radiation pattern 100 per cent a-m at 30 cycles. Since the two pairs of towers are driven 90 deg out of phase, the resultant radiation pattern is a figure of eight rotating at a 30-cycle rate. The reference-phase field is produced by the same carrier, modulated with a voltage derived from a tone-wheel generator carried on the goniometer rotor shaft, and is radiated from the central tower. The modulating voltage corresponds to an audio signal subcarrier with 210-cycle central frequency and varying from 180 to 240 cps.

In an earlier development model, two individual basic carrier frequencies were used, each carrying its own 30-cycle modulation. The reference-phase field was produced by a carrier (say, 172 kc), modulated with 30 cycles and radiated from the center tower. The variable-phase field was produced by a carrier (say, 194 kc) feeding the rotor goniometer coil of the differential antenna system. A two-channel receiver was required to reduce the complex received field to the 30-cycle reference- and variable-phase signals, and these signals were combined as in the u-h-f omnirange receiver to produce course indications along any radial toward or away from the omnirange station.

Although the disadvantages of high atmospheric noise level and interference with reception by precipitation static are present as in the obsolescent l-f four-course range, the relatively few l-f omniranges required for large area coverage render it feasible to attempt to counteract these disadvantages through the use of much higher power and larger antenna structures. The very narrow side bands used for the directional information also afford the possibility of utilizing highly selective receivers thereby reducing the amount of noise accepted.

The problem of reduction of accuracy of the course indications, because of polarization error from the nighttime ionospheric wave, is also still present. As in the l-f directional range, polarization errors are minimized by confining the radiation to the vertical towers, and eliminating in so far as possible the radiation of horizontally polarized components from the transmission lines feeding the two pairs of towers producing the directional pattern. Since the ground will not support the transmission of horizontally polarized waves, radiation of the undesired components can occur only by way of ionospheric propagation. If present, such components (having directional characteristics differing from those of the tower pairs) result in random discursions of the figure-of-eight pattern of the reference field about its desired uniformly rotating positions. This produces variable and unpredictable errors. Although there is as yet no accumulated field experience, analytical considerations would appear to indicate some superiority for the l-f omnirange over the l-f directional type, as to the degree of freedom from polarization errors. The lower frequencies planned for operation (see Table 2) will result in an effectively lower ionospheric reflecting surface so that lower elevation angles will be involved in the ionospheric wave (assuming that multiple-loop transmissions will alternate more rapidly than the single-loop wave). The ratio of tower height to wavelength is planned to be greater (despite the lower frequencies) thereby increasing the ratio of vertically polarized to horizontally polarized components in the ionospheric wave for a given set of conditions.

24. Other Omnirange Systems. An interesting narrow-band omnirange system, proposed by J. D. Woodward, employs a nondirectional control antenna radiating at carrier frequency and a directional four-element antenna system transmitting a figure-of-eight pattern rotating at a frequency above or below the carrier frequency by a small value (600 cycles proposed). Along a selected radial from the omnirange, a monitor receiver beats the two carriers to produce a 600-cycle signal of reference phase, which is frequency-divided to 200 cycles and applied as a reference modulating signal to the carrier on the central radiator. At the air-borne receiver output, two signals are present: a 200-cycle signal of reference phase and a 600-cycle signal of variable phase. The former is frequency-multiplied to 600 cycles, and the two are compared in phase to yield the bearing with respect

to the reference radial direction. The two carriers at the transmitter must be controlled accurately to maintain exact frequency difference to the required degree of accuracy.

A microwave omnirange system, operating at 12 cm, has been demonstrated experimentally by the Sperry Gyroscope Company. The system is basically similar to the v-h-f omnirange with the variable-phase field carrying a 1-f variable-phase signal and the reference-phase field modulated by a 70-kc subcarrier which is in turn modulated by the same 1-f., but of constant phase. However, since the complete antenna structure is only $2\frac{1}{2}$ in. in diameter by 8 in. high, it may be rotated, *in toto*, to secure the desired pattern distribution. The system should be equivalent in operating characteristics to the u-h-f omnirange with perhaps a somewhat greater potential accuracy because of better control of the directional pattern.

The Sperry Gyroscope Company has also proposed an omnirange system, operating at 6 cm and employing a nondirectional and a rotating directional antenna pattern. The rotating pattern is to consist of two intersecting radiation lobes modulated at different audio frequencies and is to be rotated uniformly in azimuth at a speed of several revolutions per second by mechanically rotating the antenna and dish reflector. Alternatively, some scheme of electrical rotation (by phase shifting) may be used. The nondirectional antenna radiation is modulated at several thousand cycles per second and keyed at a rate synchronized with the rotation of the directional antenna pattern so that the keyed pulses constitute a definite azimuth reference. At the airplane, the indicator essentially measures the time interval between the arrival of the nondirectional reference pulse and that of the equisignal course of the directional pattern.

The Federal Telephone and Radio Corporation (F.T.R.) has under development an omnirange system which forms an integrated part of its proposed coordinate Navar system. The ground station transmits a nondirectional synchronizing signal from the 1,000-Mc DME ground beacon whenever the rotating directional antenna of a 3,000-Mc surveillance radar points north, and a 3,000-Mc signal whenever the antenna points to the receiving airplane. The rotational rate of the directive antenna is uniform at about 30 rpm. On the aircraft, the DME receiver picks up the north-synchronizing signal and a 3,000-Mc receiver (forming also part of an air-borne transponder unit) picks up the directional signal. The indication of bearing, based on measurement of the time between the arrival of the two sets of signals, may be displayed on a scope with either a circular or horizontal sweep, and may also directly operate a dial-type azimuth meter.

The writer has proposed an omnirange system, operating at v.h.f. and employing a nondirectional and a rotating directional pattern in combination. In one form, the directional pattern is a figure of eight rotated electrically at a 1,350-cycle rate. The nondirectional antenna is keyed with a double pulse each time a selected null of the directional antenna points north. Both the nondirectional and the directional antennas are keyed simultaneously for each revolution of the directional antenna pattern when the null lies at four bearing angles corresponding to θ , $\theta + (\pi/2)$, $\theta + \pi$, and $\theta + (3\pi/2)$, the value of θ being progressively advanced at a rate of 2 deg per revolution. The pulse width may be 1 μ sec. The directional and nondirectional antennas are phased to produce a rotating cardioid, the selected null corresponding to the zero of this cardioid. A complete sequence of bearing pulses occurs every 45 revolutions of the directional pattern, *i.e.*, 30 times per sec. The bearing pulses corresponding to $\theta = 0$ deg and 2 deg may be omitted to avoid interference with the north reference pulse. At the receiving point, the received signals may be presented on a cathode-ray tube, with a horizontal sweep rate of $\frac{1}{30}$ sec. The reference pulse may be gated to appear on a separate parallel sweep with separate controls. The cathode-ray pattern will appear as a series of 178 pulses amplitude-modulated according to a half-sine wave, with the reference north pulse somewhere along the wave. The time difference between the zero of the wave and the reference pulse gives the bearing in degrees. The measurement may be made to a few microseconds by standard techniques; thus, the potential indicated accuracy is better than 1 deg.

In another form, the rotating directional pattern comprises two figures of eight, crossed at right angles. This is readily achieved by a mechanical goniometer. The nondirectional antenna is keyed with a distinctive double pulse, 50 μ sec apart, each time a selected equisignal, formed by the two rotating figures of eight, points north. The two directional rotating patterns are keyed alternately at equispaced, 50 μ sec, intervals with the nondirectional antenna keyed in synchronism with each to form cardioid patterns. Thus, only one equisignal course is rotated. Alternate keying pulses may have different widths, say, 1 and 2 μ sec, for distinguishing purposes. The goniometer may be rotated at any desired rate, say, 1,800 rpm. At the receiving point, one form of display is on a cathode-ray tube with horizontal sweep. Two sets of pulses forming overlapping half-sine patterns

are observed. The time difference between the equisignal point and the north reference may be measured by standard techniques. Circuit gating and smoothing arrangements for providing an azimuth indicator and a zero-center course indicator also appear feasible.

Both systems, with appropriate modifications as to antenna, pulse width, and rate, etc., are applicable to v.h.f. and l.f. L-f operation affords the possibility of discriminating against sky-wave return on the basis of receiver gating, thereby minimizing polarization errors.

25. Consol (Sonne). During the first part of the war, Germany developed a l-f multiple-course radial system for use in long-distance navigation, based on the inter-

section of the two patterns shown in Fig. 5. The antenna system comprised three radiators, as shown in Fig. 14a, spaced 3λ apart. The antenna current in radiators *A* and *C* are one-fourth of the current in radiator *B* and, respectively, 90 deg leading and lagging in phase. This antenna combination is exactly equivalent to the two-radiator system producing the pattern of Fig. 5.

In the initial system, called Elektra, the phases of the currents *A* and *C* were successively reversed according to a dot-dash time cycle ($\frac{1}{6}$ sec for the dot and $\frac{5}{6}$ sec for the dash, see Fig. 14b), so that one of the overlapping patterns of Fig. 5 was characterized by a series of dots and the other by a series of dashes. On the equisignal courses, the dots and dashes merged into a continuous signal.

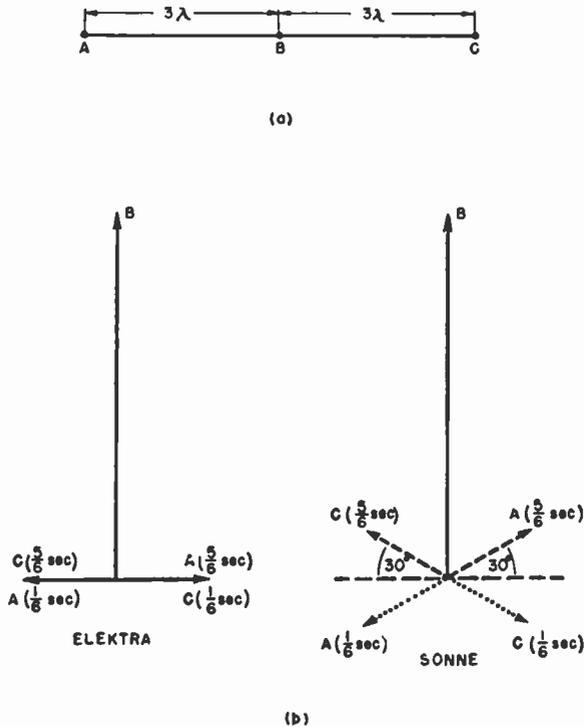


FIG. 14. Antenna spacing and phasing for Consol beacons.

The system later evolved into the Sonne, in which the phases of the currents *A* and *C* are shifted from their respective 90-deg leading and lagging values, continuously in opposite directions, during 1-min intervals, at a rate of 3 deg per sec. Simultaneously, the currents in *A* and *C* are reversed according to the dot-dash time cycle: ($\frac{1}{6}$ sec for the dot and $\frac{5}{6}$ sec for the dash, see Fig. 14b). There is thus produced a slow rotation of the equisignal lines such that at the end of the 1-min intervals the dot-and-dash patterns have become completely interchanged, and each equisignal line then occupies the position of the one adjacent to it at the beginning of the interval.

Sonne was sequenced on a 2-min base according to the following schedule: phase shifting and keying in the manner described for 60 sec; silent period for 1 sec; steady transmission from center antenna (*B*) alone for 56 sec; and silent period for 3 sec. Assume that the location of an observer is such that at the start of the cycle he hears the dash signals. Then, as the equisignal lines undergo their gradual shift, the dashes will decrease in contrast until they disappear in the equisignal. Further shifting of the

pattern is accompanied by the reception of aural dots which first increase in contrast and then decrease slightly until the end of the 60-sec period which marks the end of the phase-shifting cycle. The amount of this decrease in contrast depends upon the displacement of the observer with respect to the position of the equisignal line at the start of the phase-shifting cycle. The position of the observer within the particular sector defined by two adjacent equisignal lines is determined by a count of the dot or dash characters preceding and following the equisignal.

Since the radiated pattern produces many possible sectors, the particular sector in which the observer is located must be identified by other means. This identification is accomplished by d-f measurements on the steady nondirectional signal transmitted by the Sonne station following the phase-shifting cycle, after a 1-sec break. The steady signal is transmitted for 56 sec and is succeeded by a 3-sec break before resumption of the phase-shifting cycle. The entire cycle for obtaining bearings is thus 2 min in duration.

The Sonne system was used extensively by the Germans during the last two years of the war. The Allied Air Forces found the system reliable and also used it. The British use of Sonne transmission was usually referred to as "Consol." A number of Consol stations are in operation in Norway and Spain, and the erection of additional facilities was in the planning stage in 1946. Because of the simplicity of receiving equipment required, a modified two-antenna Consol (Sonne) system was under development by the U.S. Army Air Forces for use in the Pacific theater. The operating frequency was between 200 and 300 kc and the power in antenna B, 20 kw. To reduce potential polarization error during ionospheric propagation, unwanted horizontal radiation from the transmission lines feeding the vertical radiators was minimized by carrying the energy to the antenna towers at a low power level and then employing power amplifiers at the base of each tower.

Many qualified persons consider Consol superior to loran as a long-distance navigational aid. With the use of a simple, narrow-band, air-borne receiver, the system provides bearing accuracies within 1 to 2 deg at useful distance ranges over water up to 1,000 miles by day and 2,000 miles at night for a transmitter power of about 2 kw. The course ambiguity may be resolved by a direction finder with a rather low order of accuracy (± 10 deg). However, even at low frequencies, the use of an air-borne loop-antenna d.f. does not guarantee this order of accuracy under all conditions. It should be noted that the transmitting system may be either tone-modulated or c-w; in the latter case, the receiver and the d.f. must incorporate a beat-frequency oscillator.

One of the objections to Consol is that complete 360-deg coverage is not obtained. Another is that 2 min are required to determine a line of position and 4 min to determine a fix, *i.e.*, by bearings on two geographically separated stations. Proponents of the system point out that, with the two-antenna system, a third mast, located on a line through either one of the two and perpendicular to the base line, may be used to make the system omnidirectional. They also show that the time for taking a bearing or a fix may be reduced through the use of a simple dot-and-dash counter or printing device, with the dot-dash time cycle speeded up at the transmitter. Because Consol is a narrow-frequency system, they point to the possibility of carrying it down to very low frequencies (say, 30 kc) to secure long-range ground-wave transmission. L-f omnirange can, of course, also be used at such frequencies, but l-f loran encounters serious difficulties because of the need for pulses the very characteristics of which are broad band.

Consol charts and tables are available or in preparation for geographical use in Europe covered by existing installations. The charts show the position lines marked in degrees azimuth from the stations, and the tables provide data by means of which azimuth lines of position corresponding to given counts and d-f bearings may be evaluated.

26. Federal Telephone and Radio Corporation "Navaglobe" System. The Navaglobe system is a c-w long-wave (70-kc) omnidirectional range beacon on the ground employed with air-borne narrow-band receiving and d-f equipment. The operating

frequency was selected on the basis of exhaustive study of propagation characteristics as the frequency range best suited for reliable 1,500-mile service over water, day or night. On this frequency, investigations showed that two-tower flattop antennas with effective heights of 300 ft would require input powers from about 10 kw in northern United States to 100 kw in the tropics to give adequate signal-to-noise ratio for very narrow-band reception under the worst propagation and atmospheric conditions at distance ranges up to 1,500 miles.

The ground station for this proposed system employs three antennas located at the corners of an equilateral triangle and energized successively in pairs to give the radiation patterns of Fig. 15. A fourth nondirectional radiation is transmitted after each sequence of directional patterns for synchronizing purposes. The complete cycle of transmissions is repeated once per second. The air-borne receiver incorporates a three-coil ratio meter to measure the ratios of successive signals and utilizes these amplitude comparisons to actuate an automatic, direct-reading, azimuth meter. In addition, a shielded loop antenna provides an ADF which operates a dial indicator

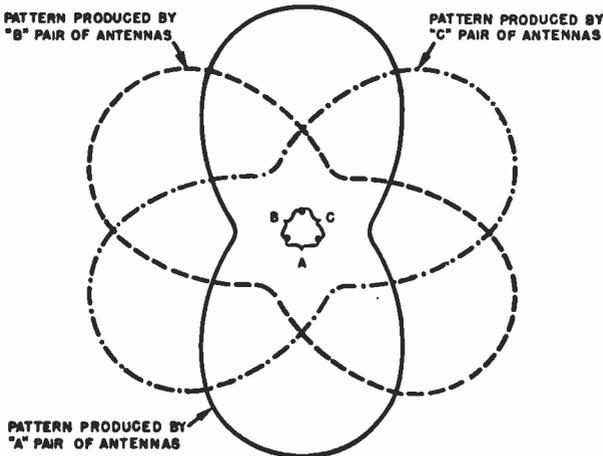


FIG. 15. Antenna patterns in horizontal plane for F. T. & R. "Navaglobe" system.

giving direct readings of the bearing of the aircraft with respect to the transmitting station. The azimuth meter readings and the d-f indications serve as useful mutually checking facilities, since the errors likely to be introduced are quite different for the two types of indication.

The F. T. & R. study showed that 60 Navaglobe stations, each providing a working radius of 1,500 miles over sea water, would suffice to cover all ocean areas, and that the entire earth's surface, exclusive of the south polar cap, could be covered by a total of 75 stations. Estimated accuracy of azimuth indication is of the order of 1 to 2 deg.

27. Air-borne Direction Finders. Various d-f systems have been devised to provide an aircraft pilot with a means of "homing" on a ground-transmitting station or of determining the azimuthal direction of arrival of the radio waves from the ground station. When used alone, air-borne direction finders provide only the heading of the aircraft with respect to the transmitting station, but when used in conjunction with the magnetic compass, they permit the determination of lines of position. The indicated accuracy of air-borne d.f. is within 3 deg, provided calibration is made on the aircraft (or type of aircraft) on which the d.f. is to be used. Absolute accuracy depends upon the propagation characteristics at the frequency used. Since the directional properties of a loop antenna are utilized almost universally in direction finders aboard aircraft for frequencies below u.h.f., the accuracies attained to date have not

proved sufficient to warrant navigational reliance solely on the basis of this aid. Experience in the v-l-f range may yield more satisfactory results. Nevertheless, navigators like d-f equipment because of the simplicity of indication, particularly as compared with the indications from other aids in current use. Accordingly, direction

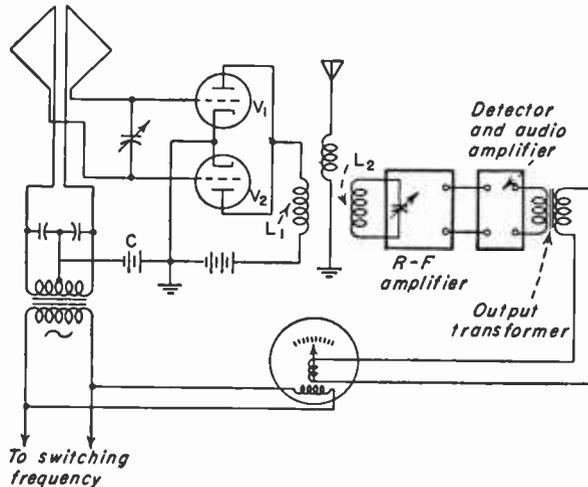


FIG. 16. Schematic circuit diagram for visual-type air-borne radio direction finder.

finders have become a standard item of equipment on all aircraft large enough to carry them in addition to the basic communication and navigational equipment.

The principle of operation of the earliest published¹ type of air-borne d.f. which is similar in most of the essential details to many of the current commercial units is shown in Fig. 16. This type is quite simple and may be used on any type of ground station.

In this arrangement the tubes V_1 and V_2 are biased to cutoff, passing current only when successive $\lambda/2$ of the switching frequency alternately make the grids less negative. The r-f voltage passed on from L_1 (in the common plate circuit of V_1 and V_2) to L_2 (connected to the input of a conventional receiving set) is then alternately reversed. Voltage from a vertical antenna is also fed into L_2 in proper phase relation so that the loop antenna voltage alternately adds to and subtracts from it. The amplified sum and difference voltage is detected and amplified and then passed through the current coil of an a-c electro-dynamometer-type instrument. The field coil is excited by the switching frequency so that the zero-center pointer is deflected to the right, say, corresponding to the additive condition of the loop and vertical antenna voltages and to the left corresponding to the subtractive condition. The polar diagram indicating the response of the antenna system for the two conditions corresponding to varying directions of the airplane with respect to the transmitting station is shown in Fig. 17. The intersection of the two cardioid patterns corresponds to the zero-center or "course" position of the indicator. Whether the airplane is flying toward or away from the ground station is readily determined by noting whether the pointer deflects to the right or left, or vice versa, as the heading of the airplane is altered to the right or left of the course.

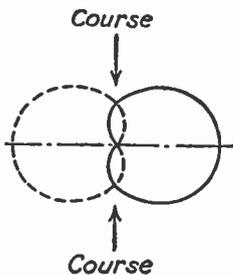


FIG. 17. Crossed cardioid patterns for visual-type finder.

Whether the airplane is flying toward or away from the ground station is readily determined by noting whether the pointer deflects to the right or left, or vice versa, as the heading of the airplane is altered to the right or left of the course.

In a modern d.f. the loop antenna is made rotatable, so that the pilot may determine "sense" by noting the relative motion of the loop and of the meter pointer. The pilot

¹ See reference to Dieckman at end of chapter.

may also determine the bearing of the ground station, by comparing the loop orientation (corresponding to zero-center meter indication) with the magnetic compass bearing.

To make full use of the possibilities of a d.f. aboard aircraft, automatic indication of the direction of the tuned-in station is required. This has been accomplished in a number of commercial units through the use of a bidirectional motor system which drives the rotating loop antenna. The motor system replaces the bilateral pointer-type indicator and is arranged to drive the loop antenna clockwise corresponding to one cardioid pattern and counterclockwise corresponding to the second cardioid pattern. A little study will show that the system is in stable equilibrium for only one course, which may be arranged to correspond to the forward direction. A bearing indicator, near the pilot, is attached to the loop antenna driving shaft by means of a flexible drive and indicates the direction of the station correctly at all times. A-v-c reception is employed to render the d-f operation fully automatic. Two automatic units of this type may be used to take continuous bearings on two stations and the resultant information used in an automatic computer to give a continuous fix in relation to the two stations.

The block diagram of a d.f. of this type is shown in Fig. 18. The amplified voltage from the loop antenna is applied to a balanced modulator which is controlled by an a-c switching voltage. The output of the balanced modulator, which consists of the two side-band frequencies (carrier-suppressed) formed by modulating the incoming carrier with the switching frequency, is combined in the antenna coupling circuit with the nondirectional antenna voltage. The resulting a-m voltage is amplified and detected by the receiver and the output signal is passed through a band-pass filter to remove all modulation frequencies

except that produced by the balanced modulation and is then further amplified. If either of the nulls of the loop antenna is pointing to the station, the resultant signal voltage output will be zero. When the loop swings through a null, the phase of the signal-voltage output reverses. This output voltage and the a-c switching voltage are applied to the two windings of a reversible two-phase motor which drives the loop (through a gear reduction). The phase shifter introduces a 90-deg phase displacement so that the currents in the two windings of the motor are 90 or 270 deg apart, depending upon the direction of the loop null, i.e., the phase of the signal output voltage. The direction and the speed of the motor thus depend upon the phase and magnitude of the signal voltage. The motor will rotate the loop to the null for which the system is in stable equilibrium. An antihunt voltage proportional to the speed of the loop of the motor is applied in series with the signal output voltage to minimize hunting of the loop about its on-course position.

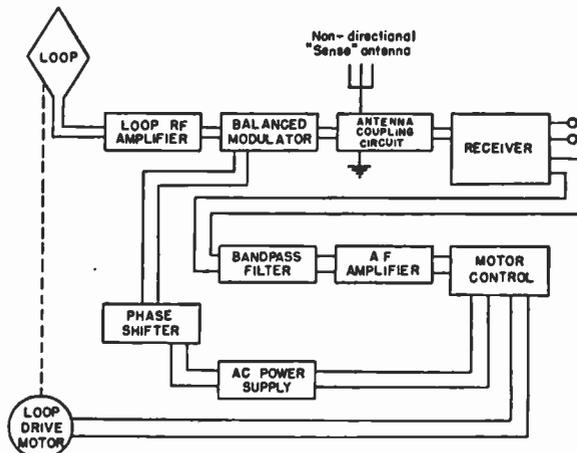


FIG. 18. Block diagram of automatic self-orienting airborne radio-direction finder. (Courtesy of Air Matériel Command.)

28. Ground-station Direction Finders. Various types of ground-station d.f. have been devised for use in conjunction with air-borne transmitters operating in the l-f, m-f, h-f, v-h-f, and u-h-f bands. Two or more d-f installations located at widely sep-

arate sites will permit a fix on the location of an aircraft. Position determination can also be accomplished by means of a single d-f operating in conjunction with range-finding equipment.

The useful range of a ground d.f. depends upon many variable factors, notably the frequency of transmission, the location and power output of the transmitter, the time of day, and the season of the year. H-f and l-f direction finders are useful at distances up to several thousand miles; whereas v-h-f and u-h-f transmissions are limited to line of sight. Present air-transport planned uses for ground-station direction finders include emergency ground aids to long-range navigation (h-f) and use at airports (v-h-f) as an auxiliary to radar surveillance.

Average bearing errors with d-f equipment are about 1.5 and 3.5 deg¹ for observations made on ground waves and on sky waves, respectively. In the case of h-f transmissions (2 to 30 Mc) the phenomenon of skip distance frequently permits satisfactory observations at great distances as well as at short distances from the transmitter, but causes poor or completely unreadable bearing indications at intermediate distances.

29. D-f Antennas. All d-f stations require a directional receiving antenna, a radio receiver, and a bearing indicator; and control centers coordinating the operation of a station network are required for the determination of fixes. Although individual bearings can be obtained in a matter of seconds, several minutes are often required in determining an accurate fix. Complexity of the equipment, which varies from lightweight portable units to large fixed installations, depends upon the nature of the service and, in particular, upon the attention to the elimination of polarization errors. The effectiveness of d-f systems may be materially increased through the addition of a panoramic adapter which gives simultaneous visual indication on a cathode-ray tube of all signals received within a given frequency band thereby assuring that any desired signal slightly off frequency will not be overlooked.

Antenna systems used with d-f equipment are necessarily directional in their characteristics but vary radically in their design. The simplest system makes use of a vertical-loop antenna supplemented by a nondirectional antenna, the combination producing a cardioid response pattern. This system eliminates the 180-deg ambiguity present when a loop is used alone. Such an antenna can be made small enough to use on portable equipment. Loop antennas give good performance on ground-wave signals but are adversely affected by reflected sky-wave signals, the latter inducing voltages in the horizontal sides of the loop and thus introducing bearing errors or instability of bearings. These polarization errors are greatly reduced by employing adequately designed Adcock antennas.

The *Adcock antenna* is a differentially connected pair of vertical radiators particularly suited for use at fixed installations. Two crossed Adcocks, associated with an inductive-type mechanical goniometer or its electronic counterpart, form the equivalent of a rotatable single Adcock. With the mechanical goniometer, a rotatable search coil is used, coupled to two stator windings crossed at right angles and each fed by one Adcock. The electronic goniometer may take a variety of forms. In one form, two balanced modulators supply a two-phase modulating voltage to the Adcock system. The two Adcock antennas feed a common receiver, but the signal voltages corresponding to the two Adcock antennas may be individually recovered in the receiver output by reference to the two phases of the special modulating (scanning) signal. In a second form, the north-south antenna is fed to the receiver through a balanced modulator operated at one a.f., and the east-west antenna is fed to the receiver through a second balanced modulator operated at a second a.f. The resultant two side bands are amplified equally by the receiver, are passed through filters to separate them, and are available as representative of the amplitude and phase of the voltages induced in the two Adcocks by the incoming wave. With either form, "sense" of the bearing may be introduced by proper combination with the amplified and detected output of a nondirectional antenna.

¹ This assumes the most modern techniques as to elimination of polarization errors.

30. D-f Indicators. Ground-station d.f. employ either aural or visual indications, or both. Aural indication of the bearing is obtained by rotating the antenna or goniometer until a null point is reached. Visual indications are presented either in the form of a minimum reading on a meter or, more generally, as deflections on the screen of a cathode-ray tube graduated in degrees around its outer edge.

When cathode-ray tube indicators are employed, the goniometer must be rotated at high speed (either mechanically or electronically) to obtain a persistent pattern on the screen. Current practice appears to favor the use of electronic scanning with all units of the d-f antenna system remaining stationary. The cathode-ray indications may take a variety of forms according to the scanning system and the manner in which the scanning voltages and the output from the d-f receiver are supplied to the cathode-ray tube. The indicator

is usually arranged to give a pulse, lobe, or line of indication on the cathode-ray tube corresponding to the direction of arrival of the received wave. Typical cathode-ray indications for a ground station d.f. are shown in Fig. 19.

The indications shown at (a) correspond to the case for the electronic scanning where the two recovered Adcock signals are applied respectively to the V and H plates of the cathode-ray tube. The voltage from the north-south antenna deflects the cathode-ray beam up and down; the voltage from the east-west antenna deflects it right and left. A straight-line indication results at an angle depending upon the relative amplitude of the two voltages and thus corresponds to the bearing of the received wave. The 180-deg ambiguity is resolved by intensity control of the beam with a properly phased voltage from a nondirectional "sense" antenna.

The indications shown at (b) correspond to the case for the rotating search coil or the two-phase balanced modulator system. The receiver output voltage corresponds to that for a rotating single Adcock and controls the intensity of the cathode-ray beam. A circle is described on the cathode-ray tube in synchronism with the position of the search coil or the phase of a rotating field set up by supplying the two-phase modulating signal to the cathode-ray plates. Presence of a signal draws this circle inward by intensity control of the cathode-ray beam so that the nulls are indicated by a lobe shaped like a propeller, *i.e.*, the difference between a circle and a figure of eight inscribed in the circle. If the circle described on the cathode-ray tube makes one revolution for each 360-deg of rotation of the equivalent rotating Adcock, two lobes 180 deg from each other will appear, corresponding to the two nulls. If one revolution of the circle corresponds to 180-deg rotation of the equivalent antenna, the two lobes will be superposed on each other.

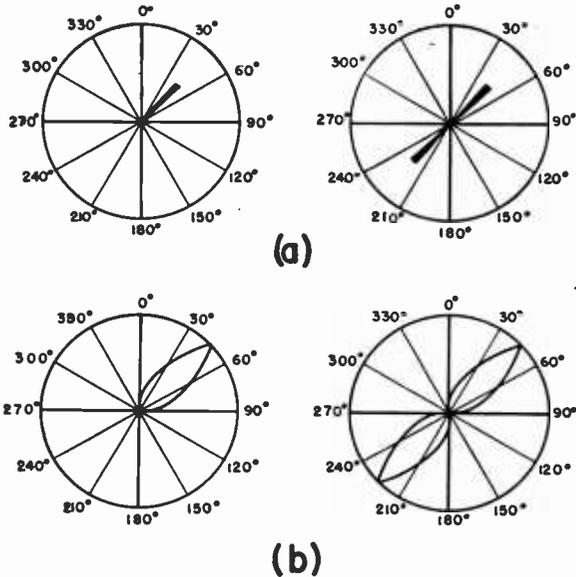


FIG. 19. Typical cathode-ray bearing indications for ground-station radio-direction finders.

A very important feature of cathode-ray bearing presentation is that visual indication is afforded of the quality of the received wave. Depending on the form of presentation, a skilled operator learns to correlate recurring modifications of the presentation with varying degrees of accuracy of the bearing. The discussion of presentation has so far related to standard forms for h-f direction finders.

Aside from polarization error any crossed Adcock system is subject to a certain amount of octantal error due to the geometry of the antenna spacing. Mechanical goniometers also introduce an octantal error due to nonsinusoidal voltage induction into the search coil from the stator windings. Both errors may be removed by calibration or by compensating distortion of the form of presentation.

HYPERBOLIC NAVIGATION SYSTEMS

31. Standard and SS Loran (1,700 to 2,000 Kc). Loran is a well-established long-distance navigational aid available over large areas of the Atlantic and Pacific, as indicated in Sec. 4. The principle of operation is illustrated by Fig. 6. A chain of stations, providing the two sets of hyperbolic position lines necessary for a fix, may consist of four stations of which two are masters and two are slaves; or one master may control two slaves, or one slave may be pulsed by two masters.

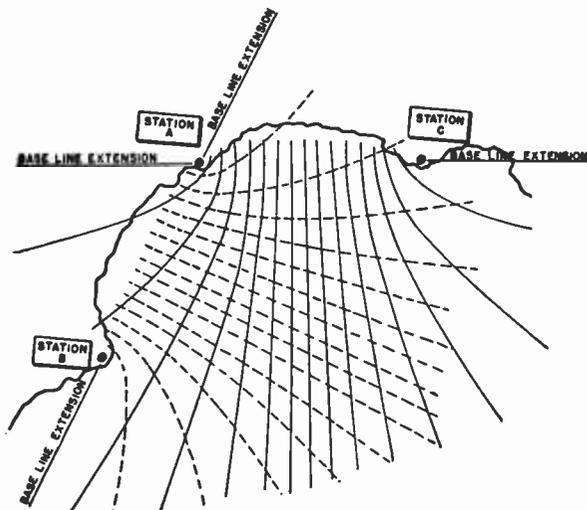


FIG. 20. Typical arrangement of three loran stations to form two loran facilities or pairs.

A 300-mile base line is conventional. Figure 20 shows a typical arrangement of three stations, with A as the master, for establishing intersecting loran position lines over an area.

Maximum accuracy is obtained when the spacing of lines is closest together and when intersecting lines are perpendicular. Since the longer the base line, the greater will be the area for which the loran hyperbolas will be approximately straight and parallel lines, two sets of loran pairs of stations, or facilities, with longer base lines and properly disposed, could be used for high accuracy coverage of an extensive area. This is accomplished in *SS loran* by relying upon sky-wave transmission for synchronization of the slave station, and base lines up to 1,200 miles are employed.

The master station of a loran pair transmits pulses (of about 80 μ sec duration and of about 70 to 100 kw of peak power) at a repetition rate which is characteristic of the particular pair. For most standard loran stations this is near 25 pps, for *SS loran* near 33½ pps. Pulses from the master station are received at the craft after an interval representing the time taken for the transmission to travel the distance from master station to craft. Pulses from the master station are also received at the slave station after some other interval which is characteristic of a given pair of stations and is pro-

vided in Sec. 4. The principle of operation is illustrated by Fig. 6. A chain of stations, providing the two sets of hyperbolic position lines necessary for a fix, may consist of four stations of which two are masters and two are slaves; or one master may control two slaves, or one slave may be pulsed by two masters. The maximum practical base-line length, *i.e.*, the distance between master and slave, is 600 miles, since reliable ground-wave reception, with adequate signal-to-noise ratio for synchronization of slave to master, is not obtainable for longer distances even over an essentially all-water path.

portional to the base line used. These received pulses cause the slave station to transmit pulses of its own at the same repetition rate. A fixed delay time is introduced at the slave station between received and transmitted pulses, for reasons explained below. The slave transmission is thus synchronized with, or locked to, the master transmission. Pulses from the slave station arrive at the craft after a time interval representing the distance from slave to craft. The craft therefore receives two series of pulses: one from master and one from slave. The time interval between the arrival of master and slave pulses is measured at the craft by means of the indicator on which the received pulses, and also suitable time-marker pips, are displayed.

The loran indicator makes use of a cathode-ray tube, on which appear two linear time-base sweeps, swept in succession and displayed one under the other. If the p.r.f. is 25 per sec, then one master pulse and one slave pulse will be displayed every 40,000 μ sec. The measurement process requires that of these two pulses, one appear on the upper part of the sweep and the other on the lower part. Therefore, the slave pulse is delayed by at least one-half of the repetition period, or 20,000 μ sec. Furthermore, to accommodate the time difference to be encountered at extreme parts of the desired coverage area, an additional "coding" delay is introduced at the slave station. These delays in the emission of the slave pulses constitute the fixed time delays mentioned above.

These and other time intervals are represented in a system of symbols which has become standard terminology in loran, as follows:

- T = indicated time difference
- T' = true time difference
- L = recurrence interval
- D = absolute delay
- δ = coding delay
- β = time taken for pulse to travel from master to slave
- C = velocity of radio propagation

All times are measured in microseconds.

Assume that a pulse is transmitted from the master station at $t = 0$. This pulse will arrive at the above station at a later time, $t = \beta$, and the slave pulse will be radiated from the slave station at a still later time, $t = \beta + \frac{L}{2} + \delta$, which is equal to the absolute delay D . At any receiving point, the true time difference between the master and slave pulses will be $T' = \beta + \frac{L}{2} + \delta \pm \frac{X}{C}$ (where X/C is the difference in actual travel time of a radio

wave from the master and slave stations to the point). However, owing to the fact that the time difference of $L/2$ is automatically taken care of by the presentation of the two pulses on the upper and lower parts of the linear time base, the indicated time difference as used by the navigator will be $T = \beta + \delta \pm X/C$.

For standard loran, β is generally 1,611 μ sec (300 miles) and δ is automatically fixed at 1,000 μ sec. This yields maximum and minimum *indicated* time differences of 1,000 and 4,222 μ sec; i.e., $T = \delta$ to $T = 2\beta + \delta$. Loran charts are prepared with hyperbolas marked every 20 μ sec (indicated time).

The time relationship between the transmitted and received pulses for a loran pair and the presentation and interpretation of the loran signals at the point of observation are indicated in Fig. 21. All loran stations received on the same frequency will produce visible pulses, all of which drift across the screen except those having the same repetition rate as the selected oscilloscope sweep rate.

In carrying out the actual determination of a fix, the operator, by means of switches, first chooses and stops a desired pair of pulses. The controls are adjusted until the master signal is viewed near the beginning of the upper or *A* trace. The slave signal then appears on the lower trace. He then obtains a coarse alignment of each pulse with respect to an adjustable pedestal, applies a faster sweep to expand the traces, shifts the pulses until they coincide in time position and, finally, obtains the actual time-difference reading (from a counter, in modern receivers) which identifies the loran line. After making an exactly similar determination on another pair of pulses, the fix

is determined by means of a map which shows the hyperbolic loran lines. Special precautions are taken at the transmitting points to warn navigators of any lack of synchronization in the emitted signals, thus ensuring the dependability of the loran service. Such warning is effected either by causing a distinct blinking of the pulses or

by periodically causing a marked change in the time difference between the two pulses.

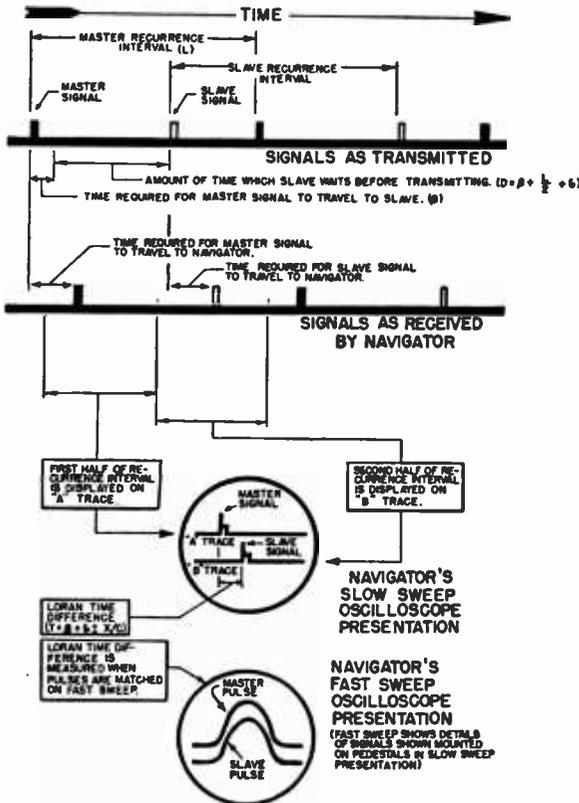


FIG. 21. Presentation and interpretation of loran signals. (Courtesy of U.S. Coast Guard.)

amplification of each of the two signals received at different ratios. This feature permits presentation to the cathode-ray indicator of signals of equal amplitude. Sufficient range of operation is provided to permit accommodation of incoming signals having a ratio of strengths as high as 500:1. Earlier equipments had an operating limit of roughly 100:1.

The indicator unit contains the necessary circuits to perform all the timing functions of the equipment with the required precision. It contains the sweep generators and the cathode-ray tube for presentation of the signals. The basic timing medium is a precision, crystal-controlled, master oscillator. The oscillator possesses high short-time stability of the order of a few parts in 10^{14} . Manual means of adjustment is provided to vary the frequency over a range of more than 200 parts in a million. This adjustment permits cycling the oscillator until the timing of the receiver-indicator is in exact step with the recurring pulses received from the transmitting stations. Through the medium of its timing circuits, the indicator provides a sequence of precise timing markers spaced at convenient intervals to facilitate measurement of time sequences with a basic accuracy of the order of $\pm 1 \mu\text{sec}$. The sweep generators provide a slow sweep as outlined in the system specification which covers the entire recurrence interval by means of a divided trace. A fast sweep lasting on the order of 200 μsec or less is provided and by means of

The radio receiver of a loran receiver-indicator is provided with four r-f channels and may be set for operation on any of the four loran frequencies. In operation, receiver tuning is fixed, and a four-position switch provides simple means of changing channels. The receiver has a sensitivity sufficiently high that a signal of approximately $10 \mu\text{v}$ delivered by the antenna to the receiver input will result in full scope deflection of the cathode-ray indicator. Receiver and indicator form an integrated unit and are not intended to function separately. Early designs incorporated a total band width of the order of 80 kc at 6 db down. Current trends are toward considerable reduction and total band widths of 40 kc or less at the same decibel limitation are contemplated. The receiver incorporates a differential amplifier which operates in synchronism with the incoming signal-recurrence frequency to permit am-

delay controls may be positioned to examine the particular section of the time base at which a signal appears. For convenience in identification and for purposes of triggering and delay measurement, a pedestal or raised rectangular pulse appears on the slow-sweep presentation in the portion covered by the fast sweep. Timing arrangements are made to permit the fast-sweep generator to fire at a predetermined point on both the upper and the lower traces. To measure loran time differences, the time base is cycled with respect to the transmissions until the master and slave signals appear at convenient points on the upper and lower traces, respectively, with adjustment being made such that the fast-sweep generator fires precisely in the region of both signals. The signals are examined on the scope operating on fast sweep, and a fine adjustment is made until the pulses are superimposed or matched with respect to time. Time-difference measurements are made by means of timing markers or by automatic counters in the more modern receivers.

Standard loran will ordinarily furnish reliable daytime service over water at distances up to 750 nautical miles. This represents the extent of the ground-wave transmission. The nighttime range is extended by means of ionospheric propagation to 1,400 nautical miles. SS loran, relying solely upon ionospheric transmission, provides a maximum range up to 1,400 nautical miles. However, areas within about 300 nautical miles of the individual stations are not covered, since this would usually involve matching of one ground wave and one sky wave.

Various errors are inherent in the different elements of the loran system, such as lack of perfect synchronization between the transmitted pulses, errors in sky-wave corrections, and errors in charts and tables. Furthermore, the area of best accuracy is where the hyperbolic lines are closest together. This occurs along the base line between a pair of transmitting stations and also in the vicinity of the perpendicular bisector of the base line. The position accuracies obtainable thus vary over a wide range. In general, near the base line, the accuracy of a loran fix is within 1,000 ft, and even at a distance of 800 miles from the transmitting stations the accuracy will usually be within 1.5 miles.

Loran charts and tables are published by the U.S. Hydrographic Office for use aboard ships and aircraft. Each chart shows lines of position for several pairs of transmitting stations, the family of lines for each pair being printed in a distinctive color. These lines are spaced at regular intervals, usually $20 \mu\text{sec}$, and every line bears an identification symbol denoting the r-f channel, the basic pulse recurrent rate (25 or $33\frac{1}{3}$ pulses per second), and the specific pulse recurrent rate (eight values differing slightly from each basic rate). The loran tables contain in tabular form essentially the same information as the charts. They indicate the intersection of lines of position, usually in steps of $20 \mu\text{sec}$, with lines of latitude and longitude, at intervals of about 1 deg.

The following features render loran an extremely valuable navigational aid: (1) long range over water, (2) accuracy of fixes comparable to that by good celestial observations, (3) fixes obtained rapidly (2 to 3 min), (4) air-borne equipment not complex and relatively easy to use, and (5) operation of the system independent of all other navigational instruments. At its present stage of development, it lacks the means for converting its data to the form needed for ready navigation, *i.e.*, distance and bearing.

32. L-f Loran (180 Kc). L-f loran was conceived during the last year of the recent war and, through the joint effort of the Signal Corps and the Radiation Laboratory of the Massachusetts Institute of Technology, was installed that year along the Eastern seaboard. The following year, the entire installation, which consisted of a chain of stations of an experimental character, was transferred to the northwest section of Canada where it served the requirements of certain military maneuvers in that area. Shortly thereafter, a joint American-Canadian project was initiated, embracing a comprehensive program of propagation studies, operational tests, and equipment development, and utilizing nine fixed monitoring stations, together with the necessary aircraft and several hundred personnel. This experimental system, which is still in operation, has come to be regarded as an indispensable aid to navigation in the Arctic. The project has now yielded a large amount of data, but thus far analyses have been based upon limited sampling. Enough information has been obtained, however, to justify

certain conclusions regarding the possibilities of l-f loran operating at a frequency of 180 kc. A system comprising four stations located at the corners of a 1,000-mile square will provide useful coverages of 1,790,000 sq miles for day and 1,370,000 sq miles for night operation, with a rms distance error not exceeding 5 miles. It has been found from experience that l-f loran provides a satisfactory navigational service over large land areas and is suited to operation both in the Arctic and in temperate climates. Furthermore, such convincing evidence has now been obtained as to the efficacy of this system, that the Special Radio Technical Division of ICAO recommended the

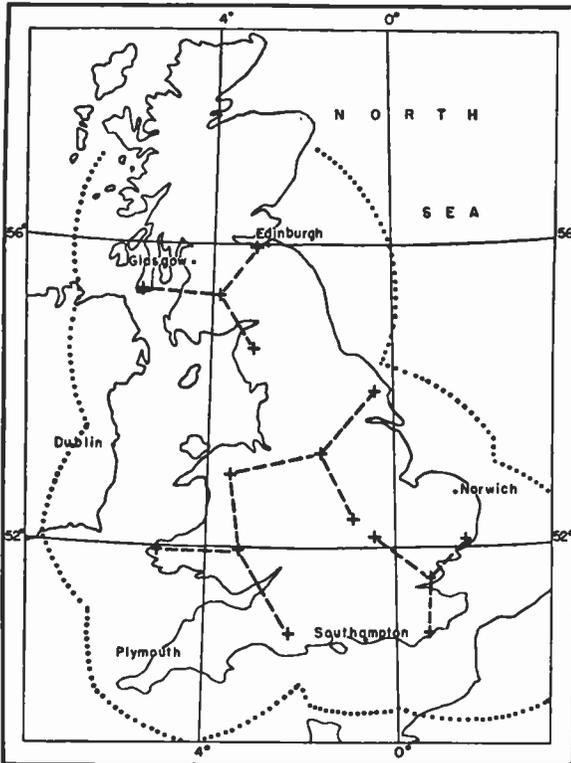


FIG. 22. Coverage of British Islands by a network of 15 Gee stations (4 chains).

prevents this cycle-matching technique from being effective at greater distances.

34. Gee Hyperbolic Position-determining System. A British air navigational system, which is known as "Gee" and which antedates the loran system, utilizes hyperbolic lines of position established by means of master and slave stations operating in the 20- to 85-Mc region. At these frequencies, the range depends upon the siting of the transmitter and the height of the aircraft, varying from somewhat more than 100 miles at an altitude of 1,000 ft to as much as 350 to 400 miles for high-flying aircraft. The Gee system, first operated early in 1942, subsequently became the standard British aircraft electronic navigational aid used during and since the war. Great Britain favors Gee over the v-h-f omnirange as a regional short-distance aid for Western Europe on the basis of better coverage for a given number of stations. With four Gee chains of four stations each (see Fig. 22) full coverage, at ground level plus 500 ft, is

the technical development, investigation, and evolution of l-f loran, as well as other long-range navigational aids, in various parts of the world. The division also recommended a study of the location of stations for world-wide coverage, with a view to coordinating the requirements of the separate regions.

33. Cycle Matching. A limited amount of experimentation has been conducted with a modification of l-f loran in which the r-f cycles themselves are matched, the pulses being used only for the purpose of identifying the proper cycles. Although difficulties were experienced with this method, cycle identification was accomplished in the case of 75 per cent of the measurements, and the average accuracy achieved was equivalent to an average error in the line of position of 160 ft at a distance of 750 miles. Sky-wave interference with the ground wave

obtained over Great Britain with moderate coverage over the surrounding waters. This coverage is based on the use of a multistack array and 240-kw power at each of the 15 stations (one station is common to two Gee chains).

All four stations of a Gee chain operate on a single frequency. The master station transmits pulses of 2 to 10 μ sec width at a pulse repetition rate of 500 per second. Two of the slave stations, after a suitable time delay, retransmit these pulses at a 250 per second rate, the two stations being synchronized to alternate pulses of the master station. The master and these two slave stations set up two intersecting sets of hyperbolic lines of position from which a fix may be determined by any receiver within the lattice. Certain intersections in this lattice are at such acute angles as to lead to considerable fix error; hence, the third slave station is so positioned as to produce a third set of lines of position which may be used in combination with either of the other two sets to cover the critical areas. For identification, the third slave station retransmits two closely spaced pulses for every third pulse of the master station.

The receiver contains a crystal oscillator which affords exact synchronism to the pulse recurrence frequency of the master station. The horizontal base line on the oscilloscope of the receiver-indicator is divided into two equal vertically displaced parallel sections. In operating the receiver, the crystal oscillator is first adjusted until the pulse appearing along the base line of the oscilloscope becomes stationary. If the pulse from the master station is designated *A*, and those from the three slave stations are designated *B*, *C*, and *D*, respectively, then the pulse line-up on the oscilloscope screen will be as follows: pulse *A* will appear at the start of each base-line section, since it is always the first to arrive and since the repetition rate is twice that of the *B* and *C* pulses. Pulses *B* and *C* will appear on the upper and lower base-line sections, respectively, and the double-*D* pulse, which has one-third the recurrence rate of the *A* pulses, will appear near the end of each base-line section. As already indicated, the *D* pulse is utilized only when the position lines determined by the *A-B* and *A-C* pulse combinations intersect at an angle that is too acute for adequate accuracy.

With the pulses in the positions described, adjustable notches in the base-line sections are fitted to the *B* and *C* pulses, and an expanded sweep is applied to magnify the portion of the display lying within the notches. By means of suitable controls, the two *A* pulses and the *B* and *C* pulses are then brought into vertical alignment. The sweep is again expanded and a last alignment is made. The procedure is to count the time-interval markers by which the *B* and *C* pulses were shifted along the base line in effecting their alignment with the *A* pulses, thus obtaining the actual differences in arrival times. Special Gee charts, showing hyperbolic lines of position for various time differences, will then permit the plotting of a fix by means of two lines of position. The accuracy of time-difference measurement with this system is better than 1 μ sec.

There is considerable controversy as to the relative merit of the Gee and the v-h-f omnirange for short-distance navigation. The former affords greater accuracy and better area coverage; the latter better instrumentation for following a fixed route if the stations are located at the route terminals. The potential merits for both systems depend upon current developments; *e.g.*, the development of a computer for flying any track within the coverage area. The technical problems involved in the development of such a computer appear considerably simpler for the omnirange system.

35. Multiple-track Radar Range. An Australian experimental short-range navigational system, known as the "multiple-track radar range," and favored by Australia as a regional aid, utilizes pulses from a master station and one slave station to establish hyperbolic lines of position. This system operates at 212 Mc and transmits 1- μ sec pulses having a peak power of 10 kw and a repetition frequency of about 5,000 pulses per second from the master station. The master station and slave station are located about 8 to 10 miles apart, and each employs an antenna consisting of three stacked vertical dipoles which produce a circular field pattern in the horizontal plane. The slave is normally quiescent and operates only after it is triggered by a pulse received from the master, retransmitting this pulse with a suitable time delay.

Discrimination among several pairs of ground stations is accomplished by the use of

different pulse-repetition rates for the various station pairs. Additional station identification is provided by Morse coding of the pulses from the master station. Distinction between the master station and the slave station is obtained by using different pulse-repetition rates at the two stations, the rate at the slave station being only one-quarter that at the master. Voice transmission over the system can be achieved at the master station by providing a second pulse which follows the main pulse by 12 μ sec and which is capable of being time-modulated.

The air-borne receiver contains a pulse-repetition frequency-selector unit which permits the selection and identification of any particular ground station. It also includes an automatic circuit which measures the time delay between master and slave pulses and provides a direct reading on a meter graduated in microseconds or in numbers identifying the individual tracks shown on a map. A pointer is provided which can be set at any chosen value of delay, and a zero-center meter indicates deviation of the aircraft to the right or to the left of the desired track. These two meters, showing the track numbers and deviations, respectively, are combined in a single crossed-pointer instrument. Flight tests with the system showed that the sensitivity of this instrument will permit the detection of off-course deflections equivalent to $\pm \frac{1}{2}$ deg, or ± 90 yd near the base line between the two ground stations where the tracks do not approach radial lines. It was found in these tests that a pilot familiar with instrument flying could maintain an aircraft on a selected track to an accuracy of better than ± 1 deg. The frequency employed limits the use of the system, in general, to line-of-sight courses. In test flights, ranges up to 100 miles have been obtained at an altitude of 5,000 ft, and ranges up to 60 miles have been obtained at 2,000 ft.

A block diagram of the air-borne receiver and meter indicator circuit is shown in Fig. 23, together with the wave form of the received signals at strategic points in the equipment. The receiver comprises a v-h-f amplifier, an oscillator-mixer, 4 i-f stages, a detector-a.v.c., a video amplifier, and a cathode follower. The output from the receiver is fed into a blocking oscillator which is paralyzed for several microseconds to remove spurious echo pulses behind the main pulse. The p-r-f selector employs a stable multivibrator which is synchronized over a narrow band of frequencies by a gate tube. The multivibrator cycle is so arranged that a trigger pulse on the control grid is effective in changing the cycle duration only if it occurs in a highly restricted fraction of the cycle. (At other times the tube is cut off by suppressor bias or low plate voltage.) Thus, only pulses whose recurrence frequency is very close to the free-running frequency of the circuit will cause synchronization.

From this multivibrator, then, are derived pulses in a fixed time relation to the master-station pulses. From the blocking oscillator is also available a family of pulses, corresponding to other master stations, other slaves, and the desired slave station. Only the pulse from the desired slave station will be in a fixed time relation to the chosen master. It is necessary to identify this pulse and measure the time interval between it and the master.

This task is accomplished by a tracking circuit, which places a pair of gates at an adjustable interval of time after the master pulse. These gates are approximately the same width as the received pulses and are spaced (on centers) by about this same time again. One is applied to the screen of one of a pair of tracking amplifiers, the other to the screen of the other member of the pair. In the absence of the gate, the screens are both cut off. The blocking oscillator pulses are applied to the control grids of both. Thus pulses will occur in the plate circuits of these tubes only when there is partial coincidence between the pulses and the gates; indeed, in one plate if the pulse is a little early, relative to the gate center, and the other if the pulse is a little late. Obviously, only a pulse in fixed time relation to the master can give a continuous output here, so that the problem is solved if the gates can be placed and held in the right position: the pulse is identified, and the time interval can be measured.

To hold the gates in place, the following circuitry is used. The pulse from the p-r-f selector fires a phantastron. This is a single-tube, one-shot multivibrator, the duration of whose cycle is controlled by the plate voltage. The wave form (Fig. 23e) results on the phantastron cathode, where the time interval $F-G$ is proportional to the plate voltage. The rise at G triggers, through a pulse sharpener, a blocking oscillator from which the two gates shown in Fig. 23f are derived. These pictures are simplified; both are actually highly damped oscillations, but only the positive swings shown are of importance.

The delayed output is obtained from one winding of the blocking oscillator transformer by simply reversing the connections to it, so that the overshoot, rather than the pulse proper, becomes the gate. As mentioned above, these gates are applied to the screens of the tracking amplifiers. The information available in the plates of these tubes is used to correct the phantastron plate voltage and thus keep the gate system centered on the slave pulse. The master-slave time delay is then measured by measuring the phantastron plate voltage. The deviation from a desired course is measured by comparing this voltage with a reference voltage.

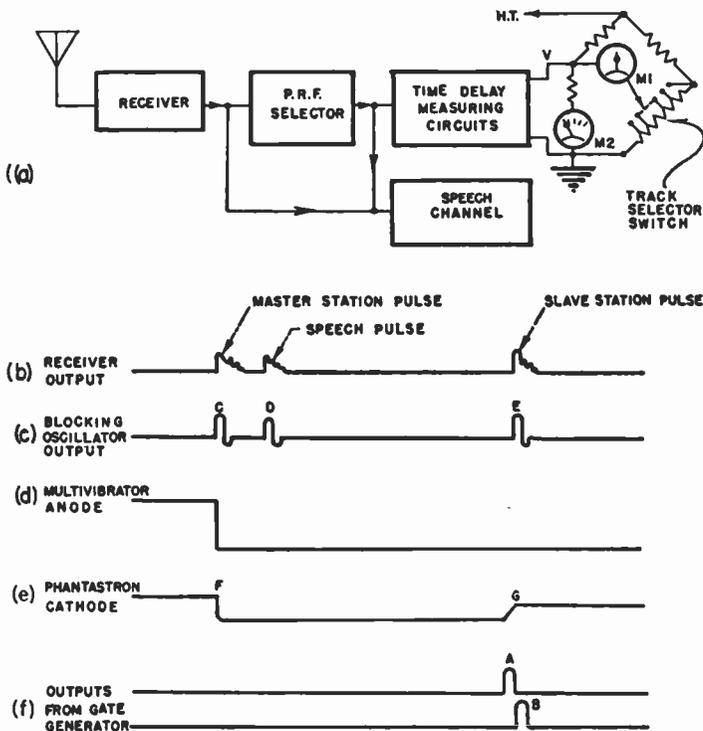


FIG. 23. Air-borne receiver and meter indicator for Australian multiple-track radar range. (Courtesy of Australian Council for Scientific and Industrial Research.)

The gate system is positioned initially by a circuit designed to sweep it slowly. If a pulse is encountered which stays in the gate for several cycles, the tracking amplifier output locks out the sweep and is thus free to track this pulse. A time constant is incorporated in the lockout so that sweeping will not recommence until failure of any pulse to appear in the gate system occurs on several consecutive cycles.

The principal advantages claimed for the multiple-track radar range are: (1) freedom from site errors, (2) ease of installation of ground stations, (3) freedom from static disturbances, and (4) provision for as many as 60 flight tracks. The principal disadvantages are: (1) the system is not completely omnidirectional, (2) conditions of terrain or other considerations may prevent the installation of the master and slave transmitting stations at the exact spacing required for establishing the desired tracks, (3) the tracks are not exactly straight lines, but, instead, have the curvature that is inherent in hyperbolic lines of position, and (4) the tracks are not uniformly spaced.

36. Decca System. A British system in the experimental stage, known as "Decca," operates in the region of 70 to 130 kc; a special version known as "v-l-f Decca" oper-

ates in the region of 10 to 20 kc. As in the case of loran, this system establishes hyperbolic lines in space by means of signals from a master station and a slave station. Unlike loran, however, Decca utilizes continuous waves rather than pulses, and phase differences in r-f signals rather than the difference in arrival times. Because of inability to obtain separate reception of two simultaneous signals of the same frequency, the transmissions from the master and slave stations are at different frequencies. These two frequencies are so related, however, that their multiplication by different small whole numbers, such as 3 and 4, will produce the same frequency; and phase comparison follows this frequency multiplication, a complete 360-deg shift occurring each time the receiving point moves from one hyperbolic line of position to the next adjacent line.

Decca employs a cumulative counter which can be set at any known point of departure and which will thereafter indicate the total number of complete phase shifts. A fix is determined by means of two intersecting lines of position established by two pairs of transmitters (two cumulative indicators may be employed to indicate the position line on each lattice). This system provides the same information as loran and is reported to be of greater accuracy than the latter. It is also easier to operate but has the disadvantage of requiring complete air-borne equipment which is not readily usable for other purposes. Matching the r-f cycles themselves makes for high resolution, thus making detectable changes in position of only a few yards.

At 70 to 100 kc, the random-phase, sky-reflected signal which appears at distances of a few hundred miles causes ambiguities, thus severely limiting the range of the system. At 10 to 20 kc, the satisfactory range should be much greater because of the increased efficiency of ground-wave propagation and the lower effective height of the supporting ionosphere layer.

37. British Post Office Position Indicator (POPI). This is another differential-phase (hyperbolic position lines) system which has received experimental consideration in England and has been tested at a frequency of about 750 kc. Four antennas, spaced rather close together and fed by two transmitters, are employed. Antennas *A*, *B*, and *C* are located on the corners of an equilateral triangle and are fed in sequence at a 50-millisecc rate (presumably in phase) from a central c-w transmitter, with a short space between each sequence of transmissions. Antenna *D*, at the center of the triangle, is fed from a second transmitter operating continuously and at a slightly different frequency from the first. Antennas *A*, *B*, and *C* set up a lattice of three sets of hyperbolic position lines. At the receiving point, the differences in phase between the transmissions from any pair of antennas are measured through the expedient of beating the successive transmissions with the continuous carrier from antenna *D*, thus preserving their phases in the resultant audio frequencies. Several methods of effecting the phase comparison of successive signals of the same frequency have been proposed.

In this system, since the antennas are relatively close together, the lines of position generated become straight radial lines within a few thousand feet of the station.

CIRCULAR NAVIGATION SYSTEMS

38. Distance-measuring Equipment (DME). As indicated in Sec. 5, the installation of DME is planned by the CAA at all v-h-f omnirange stations and at all localizer installations for ILS. ICAO specified the use of DME with both types of facility on all international air routes. Two frequency bands have been given extensive consideration for accommodating the 45 operating channels required: (1) a 40-Mc band between 200 and 300 Mc, and (2) a band of at least 250 Mc in the 1,000-Mc region. The latter band provides also for the use of DME contemplated with several integrated systems to be described in later sections. ICAO standardized on the former band on an international basis on the ground that 200-Mc DME equipment had been developed and proved reliable, but left room for the use of 1,000-Mc equipment in the United States on a regional basis.

At this writing, several equipments are under development at different frequencies. DME facilities in the Lanac and Navar systems (as well as for the CAA planned instal-

lations) contemplate the use of 1,000 Mc. A microwave system providing distance indication as part of a coordinated traffic-control plan is under development in the band 5,000 to 5,250 Mc.

CAA distance-measuring systems will utilize pulse interrogation and response, the beacon installations at omnirange stations having a range of somewhat over 100 miles and those at ILS installations a range of about 50 to 60 miles. The anticipated accuracy of measurement is 1 per cent or 1,000 ft, whichever is greater. All beacons are to be vertically polarized, are to radiate a nondirectional field pattern in the horizontal plane, and are to cover the airspace between the minimum line of sight and an altitude of 40,000 ft. The apex angle of the inverted-cone zone of silence above the beacon is not to exceed 30 deg. Beacons are to identify themselves by means of the automatic repetitive transmission of three-letter sequences in Morse code. Air-borne interrogators and beacon responders are to operate at different frequencies. Interrogators are to transmit double pulses on the beacon-receiver frequency, and provision is to be made for the optional use

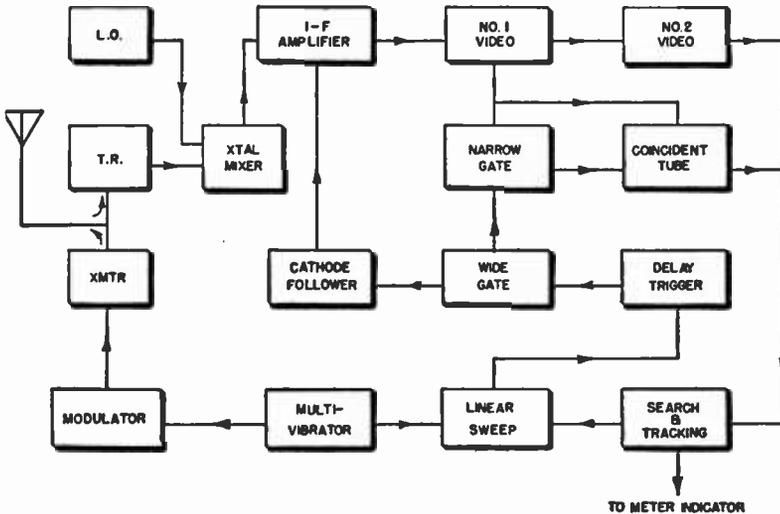


Fig. 24. Block diagram of experimental CAA air-borne DME unit (1,000 Mc).

of any one of three different pulse spacings. The beacon sensitivity is to be sufficiently high to accomplish satisfactory responder action upon receiving a signal with a peak power level of 4×10^{-12} watts. Responder signals are to be single pulses of 3 to 5 μ sec duration, and the transmitted peak powers are to exceed 5 kw and 500 watts for the navigation beacon and the landing beacon, respectively. The pulse-repetition frequency of individual interrogators is to be less than 100 cps, and all responders are to be capable of transmitting at pulse-recurrence rates up to a maximum of 10,000 cps, utilizing an automatic control system effective above 8,000 cps to limit the recurrence frequency to this desired maximum. Thus, the number of interrogations that may be handled simultaneously is of the order of 50 to 100.

A block diagram of an experimental air-borne unit is shown in Fig. 24. The equipment contains automatic distance tracking gates working in conjunction with an automatic distance search and lock-in feature. The delay between the pulse transmission and the electronic gates is made equal to the delay time T between the transmitted pulse and the received-reply pulse. This gate delay is controlled in the air-borne unit by a voltage which is increased to increase the delay. This voltage is also applied to the indicator meter. The meter then will actually indicate the delay time T and therefore can be calibrated in miles. Once a signal has been received in the electronic gates, the gates will automatically track on the signal. A wide (20 μ sec) and a narrow (10 μ sec) gate are used, their leading edges being coincident. A signal received in the wide gate will increase the voltage con-

trolling the gate delay and thus increase the delay. A signal received in the narrow gate will decrease the voltage and thus the delay. The narrow gate is given more control than the wide gate, and the signal is bracketed; therefore, regardless of the movement of the aircraft, the gates will track on the signal and the meter will continuously indicate the distance between the aircraft and the beacon.

At the beginning of operation, the delay between the transmitted pulse and the electronic gates is zero. The automatic search circuit then causes the gate-controlling voltage, and therefore the gate delay, to increase slowly (0 to approximately 1,200 μ sec corresponding to 0 to 120 miles, in 20 to 30 μ sec). When the delay is such that a beacon reply is received in the gates, the automatic search circuit will be disconnected and the gates will "lock" on the signal. The tracking will then proceed as explained. If a locked-on signal fades, the gate delay will remain substantially the same for a period of at least 30 sec. Momentary fading will occur near the end of the equipment range and during extreme banking of the aircraft. If the signal is again received within 30 sec, it will again be locked on by the gates.

Measurement of distance may be accomplished by measuring the *change in phase* of a modulation signal transmitted from point A on one carrier frequency, detected at point B, and retransmitted to point A on a different carrier frequency. The difference in phase between the transmitted and the returned modulated signals is related to the distance as follows:

$$D \text{ in meters} = \frac{\theta}{2\pi} \times \frac{3 \times 10^8}{2f \text{ (mod.)}}$$

Since phase may be measured to 1 deg, a modulation frequency of 1,500 cycles would provide unambiguous distance measurement to 100 km (approximately 60 miles) with an accuracy of better than 1,000 ft. Since the phase change is cyclical, extension of the range to 120 miles on the basis of rough knowledge of position appears entirely feasible.

The writer applied this method about a decade ago in measuring distance to a receiver-transmitter on a meteorological balloon, for use in wind-velocity determination. Checks with double-theodolite methods disclosed accuracies within a few hundred feet up to 150 miles, using a 10-kc modulation signal. This principle was used by the Germans during the war in their *Benito* blind bombing system, and in the United States it is being tested in the Aircraft Radio Laboratory (ARL) intermittent phase-comparison distance-measuring system. The method has the advantage of low transmitter power and receiver sensitivity requirements and relatively narrow-band equipment, but it requires careful attention to accidental phase shifts in sharply tuned circuits.

The Sperry Gyroscope Company has incorporated a modified method of distance measurement by phase in an integrated air-traffic-control system (to be described). The air-borne transmitter modulates the carrier successively with modulation frequencies ranging from 930 to 93,000 cycles (corresponding to the limits 100 miles and 1 mile). For any intermediate distance between 1 and 100 miles, there will be at least one frequency for which the phase shift is zero, the true distance corresponding to the lowest frequency value. If the audio oscillator can be stopped at this frequency setting, the distance from the ground station may be indicated not only on the airplane but also at the ground station.

A block diagram of the equipment proposed by Sperry is shown in Fig. 25. The modulation frequencies required by this system are obtained from an a-f generator which is driven by a servomotor. The aircraft receiver includes a phase detector which produces zero-output voltage whenever a 360-deg phase shift occurs. The output of the phase detector is used to bias a motor amplifier which, in turn, controls the operation of the servomotor. Zero-output voltage from the phase detector removes this bias and causes the motor amplifier to stop the servomotor, thus maintaining the modulating frequency at that particular value which produced the 360-deg phase shift. As the distance of the airplane from the ground station varies, the audio-generator setting will also vary automatically. As a precautionary measure, the frequency scan is repeated periodically,

starting with the lowest frequency for each scan; in this way, locking to an ambiguous signal is precluded. The ground control station obtains the same distance measurement as that obtained in the aircraft merely by measuring the critical frequency at which the range in modulation ceases, and a steady modulation obtains.

Figure 26 indicates the relationship between the distance (in miles) and the critical frequency producing a 360-deg phase shift. The higher accuracies obtainable at the shorter distances are highly desirable because of the greater density of air traffic in the vicinity of an airport.

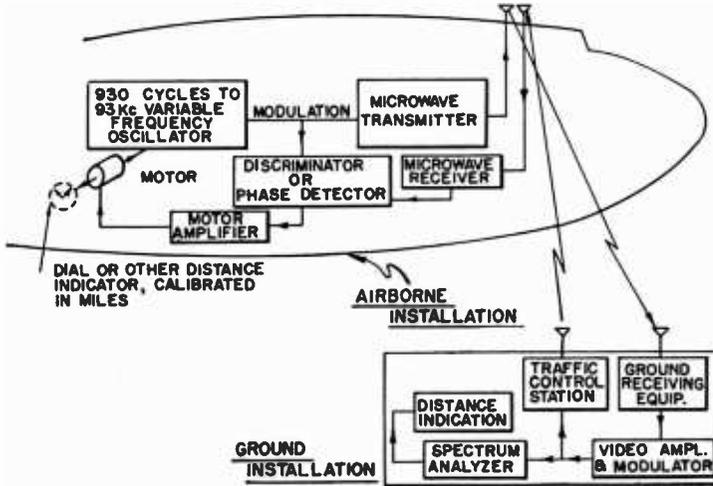


Fig. 25. Block diagram of Sperry modified phase-method of distance determination.

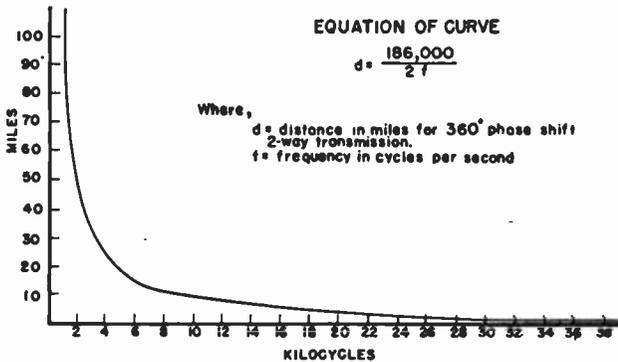


Fig. 26. Relation between modulation frequency of aircraft transmitter and distance. (Courtesy of Sperry Gyroscope Company.)

39. Shoran and Micro-H. These two systems represent practical applications of pure range measurements, *i.e.*, distance only. Both systems were used extensively during the war for precision blind bombing at distances from 100 to 200 miles. Accuracy was a first consideration in their development, overriding simplicity of equipment and man-power requirements. Accuracies of fix of the order of ± 50 ft at all usable ranges were attained.

Shoran employed two beacon responders carefully located in suitable positions (high ground) up to 150 miles apart. The airplane carried a transmitter (3 watt average

power, 210 to 320 Mc) which radiated pulses (0.5 μ sec) in alternate groups at two frequencies corresponding to the two beacon receiver settings. The air-borne receiver measured accurately the time taken for each set of pulses to travel to the corresponding beacon responder and return to the airplane. The fix was obtained on the basis of the measured slant range (corrected for atmospheric refraction) to the two beacons. Since the problem was to reach a designated target, a procedure was adopted whereby the path followed corresponded to that circular arc around one beacon station which

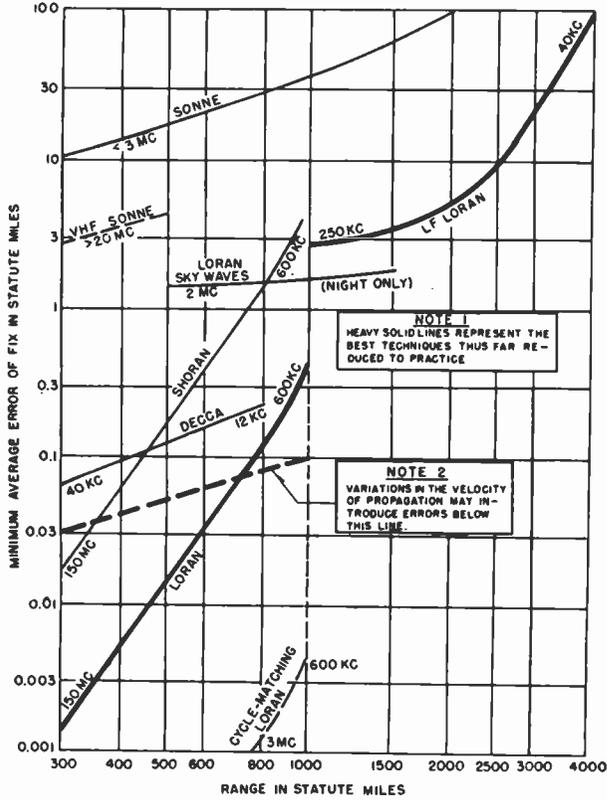


FIG. 27. Minimum errors of fix for various radio navigational systems. (Courtesy of J. A. Pierce.)

passed through the target area at the correct distance from the second beacon station. This has come to be known as a "cat-mouse" course, the cat beacon providing the guidance along the circular arc.

Micro-H, Mark II, operates on the same basic principles differing only in the complexity of the air-borne equipment. The latter comprised an air-borne radar set (AN/APS-15) with a Micro-H, Mark II, attachment. By introducing predetermined constants in switching, computer, and delay circuits, any predetermined course of a number of either cat-mouse or hyperbolic courses could be followed with respect to the two beacon stations and the bombing target.

40. Evaluations of Accuracy of the Basic Radio Navigational Systems. The foregoing description of basic radio navigational systems has included brief considerations of accuracy and of other important operational characteristics. In subsequent

sections, integrated systems will be described which employ a combination of basic systems to achieve an over-all objective, such as instrument landing of aircraft or coordinated air-traffic control. The over-all CAA plan for all classes of radio aids, which forms a master integrated system, has already been presented in its broad outlines.

A comparison of different systems involves detailed considerations beyond the scope of this chapter. Common factors must be found that will provide a satisfactory basis of comparison of widely different systems and principles. Such studies have been conducted and are in process by the government and transport agencies involved. Reports of the Cruft Laboratory, Harvard University, of the Air Matériel Command, Army Air Forces, and the COT report of ICAO represent important contributions to the present state of knowledge in this field. Summary papers on system accuracies have recently appeared.¹ General conclusions gleaned from these reports are as follows:

1. The choice of system and operating frequency for a given service represents a series of compromises as to required range and accuracy, propagation characteristics, weight of air-borne equipment, ground siting, frequency allocation (band width, etc.), and many other practical considerations.
2. Circular systems have the same position-line precision at all ranges. Hyperbolic and radial systems give a precision in line of position which varies inversely as the range, at distances greater than five times the base line. With rotating antenna systems, radial systems provide uniform precision at all azimuth angles. Hyperbolic systems yield greatest azimuthal precision in regions near the perpendicular bisector of the base line.
3. Pulse systems are not nearly so susceptible to hidden errors produced by varying propagation conditions as are C-W systems. Pulse systems require a larger band width than C-W systems and, hence, are more adaptable to high than to low frequencies. Continuous-wave systems are in general more conducive to simpler air-borne equipment, but a great deal depends upon the ingenuity of the system. Automatic indication and automatic control may be achieved with any system.

Pierce has summarized his conclusions as to useful distance range and minimum error of fix in the graphs shown in Fig. 27. He concludes that, with appropriate choice of system, accuracies of 0.1 mile or better are easily attainable to distances of about 1,000 miles and *minimum* average error of the order of 0.3 to 0.5 per cent of the distance at ranges greater than 1,000 miles. Practical average errors at ranges greater than 1,000 miles may be as much as 1 per cent of the range over a large area.

INSTRUMENT LANDING SYSTEMS

41. CAA Instrument Landing System. The basic concepts of the CAA instrument landing system (ILS) were evolved almost two decades ago by the National Bureau of Standards (operating as the Research and Development Division for the forerunner agency to the present CAA) and an integrated system was demonstrated at College Park, Md., in 1931, at the Newark, N.J., Municipal Airport in 1933, and at the Oakland, Calif., airport in 1934. The system included three basic elements to indicate the position of the landing airplane as it approached and reached the point of landing. Lateral position, given for the purpose of keeping the airplane directed toward and over the desired landing-field runway, was secured by a small, l-f (278-kc), visual-type, radio-range beacon which produced a course extending about 15 miles along the projection of the runway. Approximate distance from this transmitter was given by a distance indicator operating from the a.v.c. in the beacon receiving set. Exact longitudinal position was given by two h-f (10-Mc) fan-type marker beacons located 1,500 ft from the approach boundary of the airport and at the boundary. Vertical guidance was given by a v-h-f (91-Mc) landing beam providing a curved glide path asymptotic to the runway surface and produced by a horizontally polarized directive antenna array. Glide-path course indications were obtained by following a predetermined line of constant field intensity. Two low modulating frequencies were used for

¹ See references to J. A. Pierce and W. Q. Crichlow at end of chapter.

distinguishing the overlapping radiation patterns of the runway beacon, and the airplane-course indicator was designed to separate these frequencies by reed filters, to rectify them, and to apply the rectified voltages in phase opposition to a zero-center pointer-type microammeter. The marker beacons were of a type somewhat different from the present fan markers, but produced essentially the same vertical sheet of radiated energy. The approach marker beacon was modulated by a h-f note and the boundary marker beacon by a l-f note. These were heard in the pilot's headphones.

To simplify the indications used by the pilot, a combined instrument was employed for giving the runway-beacon and glide-path course indications. Two mutually perpendicular reference lines were provided on the face of the combined instrument, the vertical reference line corresponding to the position of the runway, and the horizontal

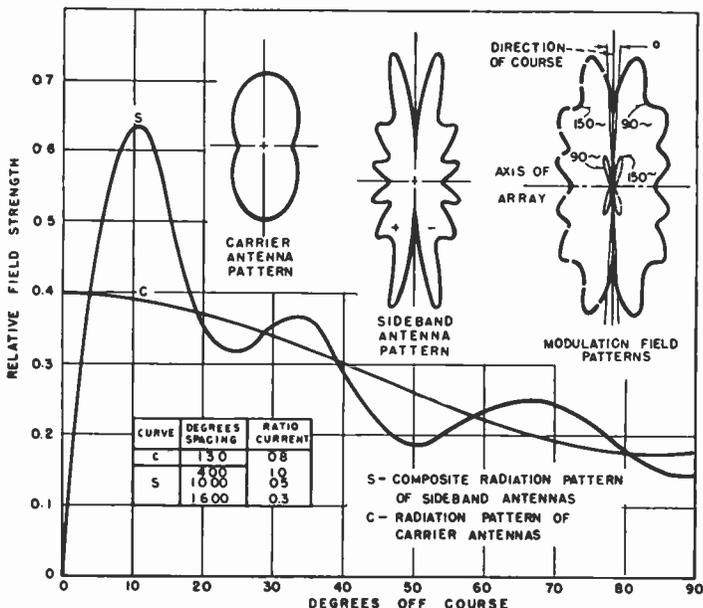


FIG. 28. Directive field pattern for equisignal localizer (ILS).

reference line to the proper landing path. The pointers of the runway-course indicator and the landing-path indicator were arranged so that they crossed each other, the position of intersection of the two pointers thus giving, through a single reading, the position of the airplane with respect to the runway and the proper landing path.

Several thousand instrument landings (hooded or in fog) were made with this system by dozens of pilots prior to 1935, including a demonstration landing in zero-ceiling zero-visibility conditions at the Newark airport following an instrument flight from College Park.

The present CAA system incorporates many improvements of the basic system evolved through the process of continuous development. The organization of equipment in the CAA system and the present status of usage were outlined in Sec. 8. A technical description of the more important elements of the system follows.

42. Runway Localizer. The equisignal-type runway localizer (LO) employs eight horizontal loop-antenna radiators (Alford type) forming a special linear array with its axis perpendicular to, and its center on, the axis of the runway. The two central radiators, connected in phase synchronism, form the carrier antenna and radiate a modulated wave,

having equal amounts of 90- and 150-cycle modulation, with maximum intensity along the runway. The remaining radiators, on either side of the carrier antenna, are connected differentially in pairs, each pair symmetrical about the center, and are fed with (carrier-suppressed) side-band energy of the same modulation frequencies but in phase quadrature with the side bands of the carrier antenna. The side-band field from the radiators on one side of the carrier antenna will be in phase synchronism with the side-band field from the carrier antenna; likewise, the side-band field from the radiators on

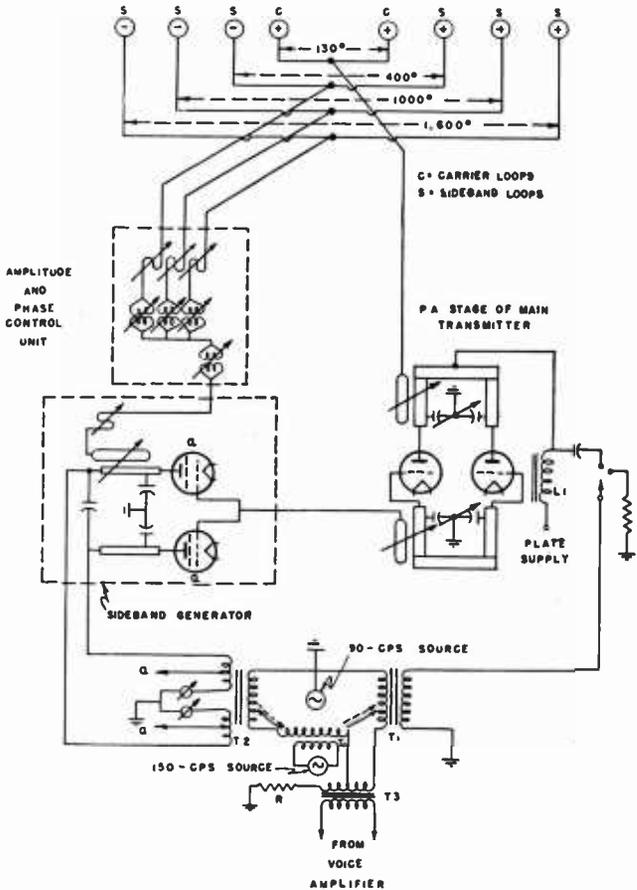


Fig. 29. Transmitting circuit for equisignal localizer (ILS).

the other side of the carrier antenna will be in phase opposition to the side-band field from the carrier antenna. For a given spacing between the radiators (130 deg for the carrier pair and 400, 1,000, and 1,600 deg, respectively, for the three symmetrical side-band pairs) a suitable distribution of side-band energy in the four pairs, expressed as a current ratio of 0.8:1.0:0.5:0.3, will yield the field patterns of Fig. 28. It will be observed that for the resultant modulation field pattern, the modulation frequency is virtually wholly 90 cycles on one side of the runway course and 150 cycles on the other side.

The runway localizer transmitting system is shown in Fig. 29.

Two main sources of r-f power feed the antenna array, one supplying modulated carrier to the carrier pair of loops (c) and one supplying pure side-band energy to the side-band pairs (s). Modulated carrier power is obtained from the output tank of a 200-watt trans-

mitter and carried directly to the antenna by means of a 70-ohm coaxial line. The transmitter is crystal-controlled, and the power amplifier stage is plate-modulated by the 90- and 150-cps signals from the motor-alternator unit. Pure carrier from the grid circuit of this amplifier is used to excite a side-band generator, the plate circuit of which is fed by the modulation voltage. The division of energy among the three side-band pairs is achieved by the fixed system of $\lambda/4$ transmission-line transformers requiring no field adjustments. The loads (R_o) are all equal to each other and are essentially 70 ohms resistive.

The phasing, which determines on which side of the course the main lobe of each modulation pattern occurs, is accomplished in a special audio hybrid circuit. The 90- and the 150-cps sources are connected into a balanced center-tapped transformer. The center-tapped winding is loaded symmetrically by the side-band generator plate circuit, and the main transmitter PA plate circuit. Under these conditions, the 90-cps components in these loads are in phase with each other, and the 150-cps components are out of phase. This same relationship (or the reverse, depending on the polarities of the coupling transformers) persists up to and including the antenna array, thus producing modulation field patterns similar to that in Fig. 28. In order that these phase relationships be obtained, it is necessary that the loads presented by the two couplings, T_1 and T_2 , be purely resistive and equal. In the actual equipment, the phase angles of the loads presented by T_1 and T_2 are of the order of 2 or 3 deg.

The 200-watt carrier is modulated 40 per cent on course by the 90- and 150-cps signals. The voice modulation utilizes the remaining unmodulated carrier. The modulation percentage observed off course will in general differ from that on course and will vary with azimuth due to the radiation pattern of the array. Voice signals are introduced by T_2 also having a center-tapped secondary forming a second balanced circuit. In this manner, the 90- and 150-cps signals are kept out of the voice equipment. R is used to maintain balance on T_2 . The voice channel in both transmitting and receiving equipment has a 1-f limit of about 200 cps to avoid interference with the basic localizer frequencies. It should be noted that voice modulation is radiated by the carrier antenna only.

A phase-comparison type of runway localizer has been developed by CAA and is preferred for regional use in the United States. This system provides for course indications by means of a comparison between two radiated r-f field patterns: one, highly directional; the other, essentially nondirectional. The nondirectional or reference pattern is modulated at a frequency of 10 kc which in turn is frequency-modulated by a 30-cycle signal. The second pattern contains a very sharp null in the direction of the course and represents side-band energy of 30 cycles, produced by a.m. of the carrier by a 30-cycle signal. Phase reversal of the modulation occurs at the null point. One side of the directional field pattern will thus carry modulation which is in phase with the modulation of the reference pattern; the other side will carry modulation which is out of phase with that of the reference pattern. When an aircraft flies along the null in the directional field pattern, no signal is received for comparison with the reference signal produced by the nondirectional field, and the receiver indicator in the aircraft therefore gives an on-course indication. On the other hand, there is a rapid increase in signal strength on either side of the null, which will produce a definite left or right off-course indication whenever the aircraft deviates from the on-course position.

The air-borne receiving equipment required with the two types of localizers was described in Sec. 22.

Course clearance and course sharpness are two of several characteristics that define the operating performance of a runway localizer system.

The ratio of amplitude of the two modulation frequencies in any direction (in db) is defined as "clearance." The clearance in a direction $1\frac{1}{2}$ deg off course is defined as "r-f course sharpness." Along the course, the clearance is obviously zero. Course width is the zone within which the clearance is less than some maximum, arbitrarily chosen. Multiple courses are said to exist if the clearance is zero in more than one direction. Bends are said to exist if the direction of zero clearance is a function of the distance from the antenna array.

Another form of course sharpness depends on the receiving equipment as well as the radiated pattern. This may be termed "a-f sharpness" and is observed by the pilot on his indicating instrument. This is a differential instrument with deflections proportional

to the differences of the two modulation voltages: $D = K(C_{90} - C_{150})$, where K is a function of the instrument itself. D depends not only on the ratio C_{90}/C_{150} (the r-f course sharpness) but also on the absolute values of these voltages, so that the deflection for a given value of C_{90}/C_{150} can be varied by adjusting the audio gain of the receiver. This is in fact done to standardize aircraft equipment.

In operation, it is important that the clearance in all directions (except on course) be adequate so as not to produce multiples or even the impression that a course is being approached. The importance of proper course alignment is obvious, and the maintenance of proper course width is equally important. These factors are, therefore, monitored continuously.

43. Glide Path. The original constant-intensity type of glide path, despite several ingenious later modifications, had one or more of the following defects to an undesirable degree:

1. Lack of range, or inability to produce a straight-line path more than a few miles in length.
2. Lack of uniformity in the glide-path angle from one airport to another and from one aircraft to another, due to differences in transmitted energies or receiver sensitivities.
3. Change of the glide-path angle depending upon the altitude of the airplane with respect to the transmitting antenna.

The equisignal type of glide path used in the present CAA system obviates these difficulties.

The overlapping r-f fields which produce the glide path are established by means of two separate antennas: one close to the ground and one directly above it and at several

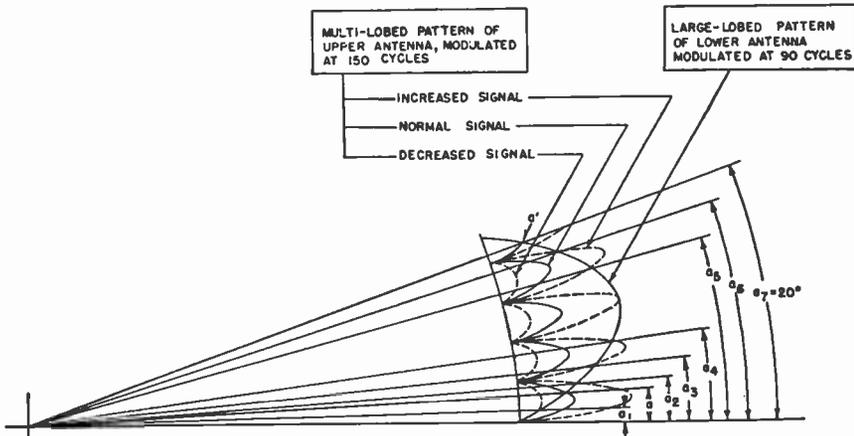


FIG. 30. Method of producing an equisignal glide path by means of intersecting field patterns radiated from two antennas at specified different heights above the ground (ILS).

times the elevation. With an r-f carrier frequency of 335 Mc, antenna heights of 5 ft 9 in. and 27 ft 6 in. will produce a glide path angle as low as 2 deg.

With this arrangement, the relationship of the multilobed patterns produced by the two antennas is as shown in Fig. 30. The pattern radiated by the upper antenna has several narrow lobes within the space occupied by the first wide lobe produced by the lower antenna. Inasmuch as the glide path is determined by the intersection of the upper and lower patterns, which are modulated by 150- and 90-cycle signals, respectively, only the lowest of the narrow lobes can be permitted to intersect the wider lobe of the lower pattern; otherwise, multiple courses would be produced. Avoidance of other intersections is achieved by properly limiting the amplitude of the field from the upper antenna compared with that from the lower antenna. On the other hand, too small a signal from the upper antenna results in a somewhat indefinite intersection, as

shown in Fig. 30. Even with the intermediate value shown, a false course is produced at the point a' , *i.e.*, 20 deg. It has been found, however, that if the ratio between the angles from the horizontal to point a' and to point a , respectively, is at least 6, no confusion is encountered.

The foregoing analysis has treated the glide path as though it extended from the region on the ground directly below the two antennas, the equisignal area produced being in the shape of an inverted cone. Actually, however, the glide-path equipment is located at a distance of 400 ft \pm 150 ft from the center line of the runway, to avoid the equipment's constituting an obstruction in the lower part of the glide path. With this arrangement, the equisignal path along the runway takes the form of a hyperbola, *i.e.*, the section produced when a cone is cut by a vertical plane other than that passing through its apex. Since the object of the installation is to provide a glide path which is substantially straight throughout its length, it is necessary to counteract the effect of the equipment being located to one side of the runway centerline. As shown previously, the glide path can be changed by changing the ratio of the amplitudes of the field patterns produced by the two antennas. Furthermore, when viewed from the glide-path station, the aircraft, in its descent down the glide path to the point of contact, subtends a considerable horizontal angle. By employing slightly different directional characteristics in the horizontal plane for the two transmitting antennas, the amplitude ratio between the upper and lower antenna patterns is made to change, with distance from the station, by the amount required to produce a straight-line glide path. The desired effect is obtained by orienting the upper antenna so that its direction of maximum radiation is displaced 12 deg from that of the lower antenna which is parallel with the runway. The disadvantages previously mentioned in changing the amplitude ratio can be tolerated under these circumstances over the short portion of the path involved.

The lower antenna consists of a half loop used in conjunction with a reflecting screen, the combination being the equivalent of a full loop. This antenna produces an essentially circular horizontal radiation pattern. The upper antenna is in the form of duplicate V-shaped horizontal dipoles mounted directly above each other at a spacing of $\lambda/2$. Both the upper and the lower antennas are clamped to a 3-in. aluminum mast which is mounted on and guyed to a small enclosure housing a 335-Mc transmitter, a power supply, and a mechanical modulator. The latter consists of a 3,600-rpm synchronous motor driving two metal paddle wheels having 3 and 5 paddles, respectively, which detune associated resonant sections of transmission lines coupled to the respective antennas and thus effect 100 per cent modulation at 90 and 150 cps.

The air-borne receiving equipment consists of a dipole antenna, mounted as a unit with the localizer antenna, and a crystal-controlled superheterodyne receiver operating at 322.6, 333.8, and 335 Mc. The meter circuit consists of 90- and 150-cycle filters and associated rectifiers. These feed the element operating the horizontal pointer of the combined instrument to indicate relative changes between the 90- and 150-cycle signals. The d-c output signals are combined in opposition, so that equal amounts of 90- and 150-cycle voltage balance the indicator and define the glide path.

In flying the glide path, the airplane approaches within 400 ft of the transmitting station, and the convergence of the patterns at these short distances causes the path to become extremely sharp and sensitive to minor displacements of the airplane above or below it. Two methods have been employed to reduce this undesirable effect. An a-v-c system has been provided for the aircraft receiver, and an additional transmitter and antenna system has been installed at the glide path station. The a-v-c system acts on the first audio tube of the receiver to decrease the audio input to the meter circuit approximately linearly with increase in signal strength, thus tending to offset the increase in sharpness of the path as the aircraft approaches the point of contact. The extra transmitter operates at the same nominal frequency as the main glide path transmitter and radiates an unmodulated carrier from a directive antenna array on the same mast. Since the array is directed across the path, its signal is received in increasing magnitude as the airplane nears the runway. The unmodulated carrier desensitizes the receiver and reduces path sharpness (or produces "course softening").

Combined Instrument. The indications afforded by the combined (cross-pointer) instrument are illustrated in Fig. 31, which is self-explanatory. Newer models of this

instrument have a flag alarm feature to give warning of loss of signal of either the localizer or glide path facility.

44. Marker Beacons. The final element of the instrument landing system is a series of marker beacons as outlined in Sec. 8. The antenna for the 75-Mc markers consists of a linear horizontal radiator, 1λ long, and supported above

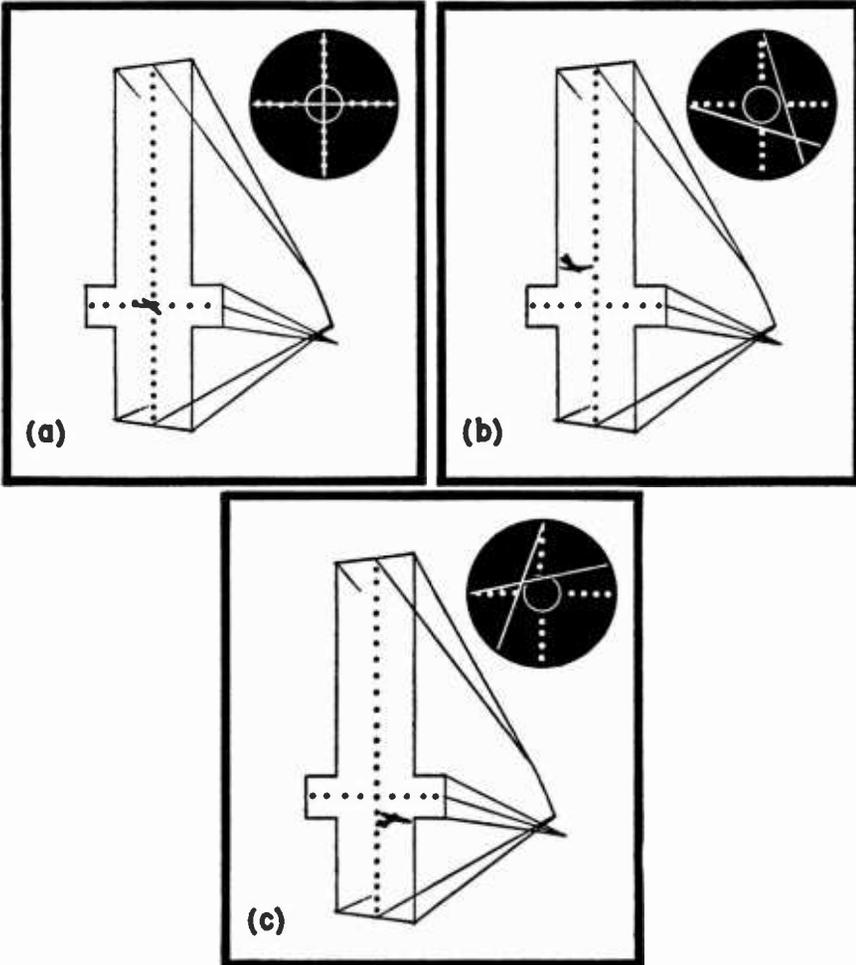


FIG. 31. Combined instrument indications for radio landing system (ILS). (a) Airplane on runway localizer and glide-path courses; (b) airplane too high and to the left; (c) airplane too low and to the right.

a counterpoise at a height of 47 deg for the boundary marker, and of 90 deg for the middle and outer markers. The marker pattern is thus broad in a plane at right angles to the runway but narrow in the direction of the runway. The marker-beacon transmitters and the air-borne receiver are crystal-controlled. The l-f homing markers (compass locators) in combination with an air-borne direction finder provide auxiliary position checks with an added facility for flying a "holding" pattern while waiting for

permission to start the landing descent. A 75-Mc "holding" marker located 10 miles from the airport assists in this holding procedure. It is the purpose of all these markers to define strategic distances from the approach end of the airport runway.

45. Automatic Approach Control. Equipment is now available commercially whereby the output of the localizer and glide-path receivers in the aircraft may be used for actuating a gyroscopic mechanism which, in turn, controls the operation of the turn and elevator controls of the aircraft. In this way, once the aircraft has come within range of the ILS, the pilot may switch from manual to automatic control, and the aircraft effects a landing automatically. The automatic technique is carried to the point of providing the necessary braking after contact with the airport runway.

A block diagram of one form of automatic approach control, developed by the Sperry Gyroscope Company, as a standard accessory to its Model A-12 Gyropilot, is shown

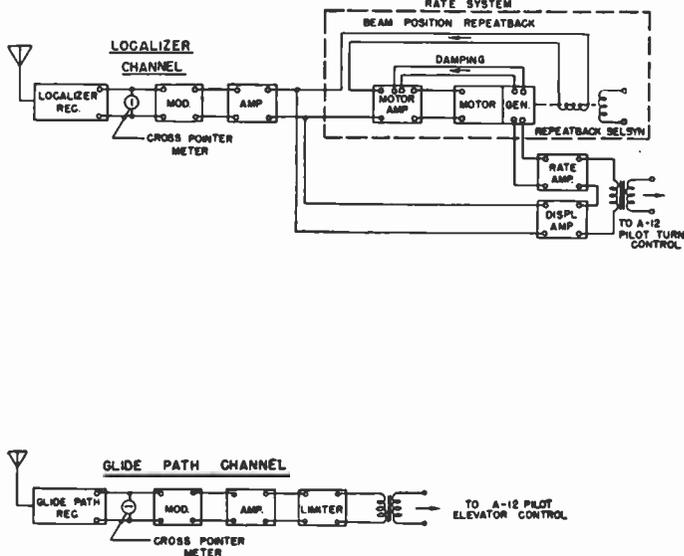


FIG. 32. Block diagram of Sperry automatic-approach control system.

in Fig. 32. To make successful automatic approaches and landings, given a suitable gyropilot, it is necessary to interpret the ILS signals into "turn" and "pitch" information for the gyropilot.

Glide-path Control. In the A-12 gyropilot, the pitch angle of the airplane is changed by adding a signal into the elevator control loop. This is normally done by the pitch knob on the flight controller. When the knob is rotated forward, a voltage is applied to the elevator amplifier which causes the airplane to nose down. When the knob is rotated back, a voltage of opposite phase is applied, which causes the airplane to nose up.

When the control is switched from pedestal controller to glide path, the gyropilot responds in the same manner. When the airplane is above the glide path, a fly-down signal is obtained. This signal is modulated and applied to the elevator amplifier to cause the airplane to nose down. When the airplane is below the glide path, a fly-up signal is obtained.

Since the glide-path receiver output is essentially d.c., it is necessary to convert the signal to a.c. A copper-oxide ring modulator is used for this purpose. This modulated signal is then amplified to get sufficient sensitivity for good control. A limiter built into the amplifier restricts pitch control to ± 6 deg. The output of this converter then is a 400-cps a-c signal whose magnitude is proportional to the glide-path signal and whose

phase reverses as the needle goes through zero. Both the sensitivity and limit can be adjusted to suit a particular airplane's requirements.

Localizer Control. Localizer control is somewhat more complex than glide-path control because the airplane is not usually on the beam when engaged, and also because turns require control of both rudder and aileron. The A-12 gyropilot contains the mechanism for turning and banking the airplane in response to voltages from either the pedestal controller or localizer receiver. The rate of turn is proportional to the voltage input, and the direction of turn changes when the input voltage phase reverses.

To prevent hunting and to have the heading of the airplane such that the ground track will coincide with the localizer beam, some method of anticipating the beam must be used. The method used is to measure the rate of change of beam signal. The heading of the airplane with respect to the localizer beam determines the crossbeam velocity of the airplane and therefore the rate of change of beam signal. Measurement of rate of change of beam signal then is used to provide damping in the control characteristic which causes the airplane to stay on the beam without hunting.

Measurement of rate of change of localizer beam signal is made by a follow-up system where a selsyn is positioned according to beam signal strength. The speed of the follow-up motor is the rate at which the beam signal is changing. A small eddy-current generator is built into the motor and generates a 400-cycle voltage proportional to the motor speed.

The modulated and amplified localizer signal makes the motor run until the output of the selsyn is equal and opposite to the motor amplifier input voltage. If the beam signal is changing at a constant rate, the motor must run at a constant rate to make the selsyn follow the changing input signal and the output of the generator is the rate of change of beam signal. If the rate of change of input signal is doubled, the motor must then run twice as fast to keep the selsyn following the changing signal, and as the motor is running twice as fast the generator output will then be doubled, indicating the increased rate of change of beam signal.

The off-course signal will cause the airplane to turn toward the beam while the rate signal is sensed to turn the airplane away from the beam. If the airplane is headed toward the beam, the off-course signal gradually decreases to a point where the rate signal will be equal and then greater than the off-course signal. Because of this, the airplane's rate of turn will decrease, and actually reverse, as it approaches the beam.

By adding beam departure signal and rate of change of beam signal in the proper ratio and feeding the resultant voltage into the turn control of the A-12 gyropilot, the airplane will turn and stabilize on the beam without hunting and with no cross-wind error.

The vertical gyro in the A-12 gyropilot is maintained vertical by means of a pendulous erection system, and during turns this erection mechanism must be turned off if the gyro is not to be precessed from the vertical. For similar reasons the leveling system in the gyrocompass control must be turned off during turns. Since the rate of turn is proportional to the output of the localizer channel, it is possible to use this voltage to distinguish between straight and turning flight. The erection cutoff section receives part of the turn signal, and after being amplified, operates a relay which provides the necessary action to cut off the erection and leveling systems.

46. Ground-controlled-approach System. One of the most important radar applications to come out of the war, this system (GCA), when used alone, places the entire responsibility for aircraft landings upon the ground controllers. So effective and relatively simple was the GCA technique that, during the latter stages of the war, aircraft were landed under conditions of zero visibility without any previous pilot training and without any special air-borne equipment. The plans for use of this system in civil air transport as a supplementary facility to ILS require certain modifications in the equipment and in its use, as detailed in Sec. 9. The following technical details relate to the wartime equipment, since this serves well enough to illustrate the basic principles.

The complete system is installed in a single van-type vehicle and consists of two separate radar sets operating at 3,000 and 10,000 Mc, respectively, together with the necessary power equipment and other communication facilities. The 3,000-Mc radar constitutes a search set which utilizes a continuously rotating antenna, installed on the roof of the van-type trailer. Duplicate, plan-position, cathode-ray indicators are employed, with range scales of 7.5, 15, and 30 miles. The 10,000-Mc radar constitutes a precision system which provides accurate elevation and azimuth information with

respect to the airport runway on separate precision indicators. Each precision indicator utilizes two cathode-ray tubes, one covering a 10-mile range, which is the limit of the precision system, and the other a 2-mile range.

The search system is used for locating all aircraft in the vicinity and directing them to the correct position for landing. The search radar transmitter operates with a peak r-f power output of 80 kw. The pulse-repetition rate is 2,000 pulses per second with a pulse duration of 0.5 μ sec. The energy is transmitted to the antenna by a wave guide and coaxial line. The antenna reflector is a cylindrical parabola and is fed by two dipoles located in the wave guide in such a manner as to produce a "coscant squared pattern," approximately 7 deg wide in azimuth, and 12 deg in a vertical plane. At the present time this type of antenna is considered the best available. It eliminates a large portion of ground clutter so objectionable on other radars and gives a wide-angle vertical pattern necessary for detection of aircraft flying at high altitudes close to the airport. The antenna rotates at 30 rpm, thus presenting a constant picture on the 7-in. cathode-ray tube indicator, and enables the operator to track the flight of the aircraft very closely. The cathode-ray tube picture is projected through a mirror arrangement so that it is superimposed on a map of the surrounding area. Thus the operator is able to tell the exact location of any particular aircraft and, accordingly, can instruct the pilot to avoid any obstacles in his path.

The antenna beams of the precision system are lined up carefully with the runway and sweep through narrow angles; 20 deg horizontally for the azimuth antenna pattern and 7 deg vertically for the elevation antenna pattern. The precision radar transmitter operates with a peak r-f power of 15 kw. The pulse-repetition rate is 2,000 pulses per second with a pulse duration of 0.5 μ sec. The energy is applied alternately to the azimuth and elevation antennas by a motor-driven r-f switch. A blanker commutator mounted on the same shaft causes the output of the receiver to appear alternately on the azimuth and elevation indicators. Both the azimuth and elevation antennas scan electrically, two to four times per second, by means of a system which mechanically varies the width of the wave guide feeding the transmitting dipole stacks. This scanning mechanism is geared to the commutator shaft and thus is synchronized with the operation of the r-f switch and blanking commutator. The azimuth antenna pattern is 0.6 deg wide in azimuth and 1.5 deg in elevation. The elevation antenna is 0.4 deg wide in elevation and 3 deg in azimuth. The precision of indication on the 7-in. cathode-ray tube indicators is thus sufficient to estimate a landing airplane's deviation from the proper glide path to approximately 5 ft, at ranges of 1 mile or less.

Reliability of this facility is increased through the duplication of most of the essential parts. With the exception of the antennas and the cathode-ray tubes, all circuits are in duplicate and may be switched instantly by a master switch. The GCA radio communication equipment provides for two-way contact with aircraft on 12 channels in the h-f band and on 9 channels in the v-h-f band. An additional receiver is used to monitor the control-tower frequency.

The nature of the information appearing on the precision elevation and azimuth indicators of the wartime equipment is shown in Fig. 33. The accuracy of the GCA system in maintaining control of the aircraft along the approach and glide paths is ± 200 ft in range, ± 0.25 deg in azimuth, and ± 0.125 deg in elevation. The accuracy of the search equipment is ± 0.75 mile in range at a distance of 30 miles.

The manner in which such equipment may be utilized in assisting a pilot following the ILS system is self-evident. Assuming that the information is available in the airport control tower on a remotely indicating three-dimensional scope, the controller is enabled to keep track of all landing operations and is in a position to provide emergency assistance, particularly for airplanes having faulty or nonoperating equipment.

To determine what features are necessary in a practical radar system for air traffic control, CAA conducted extensive tests of a number of military radar sets, including the GCA search system. Quoting from a CAA report:

This radar was expressly designed for air traffic control and contains many of the desirable features that are lacking in the equipment previously mentioned. The antenna rota-

tion of 30 rpm has met with approval from all operators and the sharp definition made possible by the short pulse width and carefully designed receivers constitutes a definite improvement over the other types. The high pulse rate compensates for the high antenna speed, so that signals are not lacking in strength. The antenna itself is reasonably small and light and the cosecant squared pattern is far superior to the other types tested. This radar detected almost 100 per cent of the aircraft in the vicinity. The blind spot, or cone of silence, directly above the antenna is considerably smaller than with other types of antennas, and ground clutter is also reduced. The projection of the cathode-ray tube

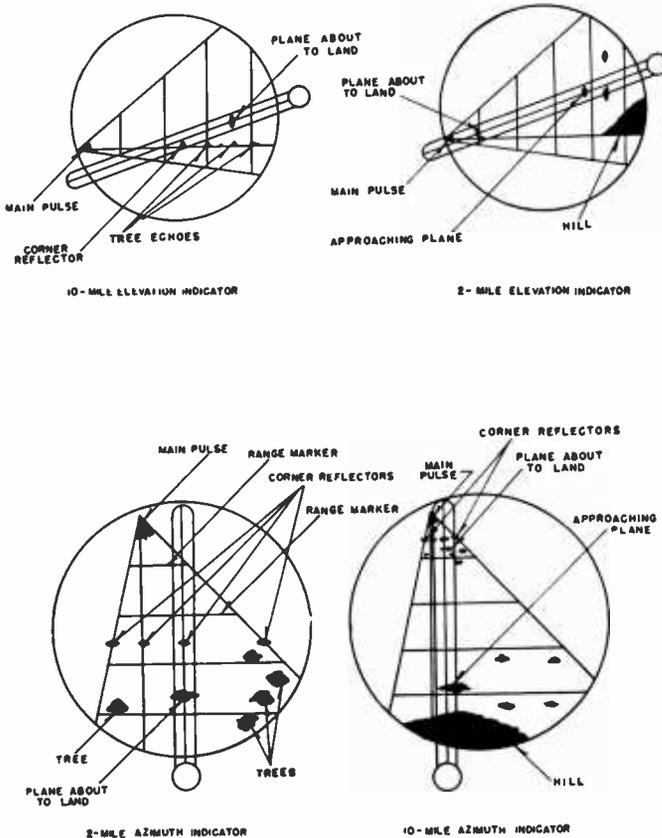


FIG. 33. Information displayed to operators of precision indicators of GCA "talk-down" radar landing system. (Courtesy of Air Materiel Command.)

picture upon the map has proved to be desirable, and the provision of sensitivity time-control circuits and fast time constants has improved the presentation.

47. Sperry Microwave Instrument Landing System. This system is based on a careful appraisal of the requirements of an instrument landing of aircraft for either manual or automatic approach and landing. Basic considerations included the form of guidance, the nature of the presentation, and the required absolute accuracies of the indications. Sperry concluded that a modulated c-w runway-localizer and glide-path system, together with DME, was essential and that the requisite high order of precision and stability of indications pointed to the use of microwaves of 3,000 to 5,000 Mc.

Ground and air-borne equipment for the system are currently available for the lower frequency and in development for the higher.

The ground transmitting equipment consists of two independent trailer units: one housing the localizer transmitter and one housing the glide-path transmitter. The

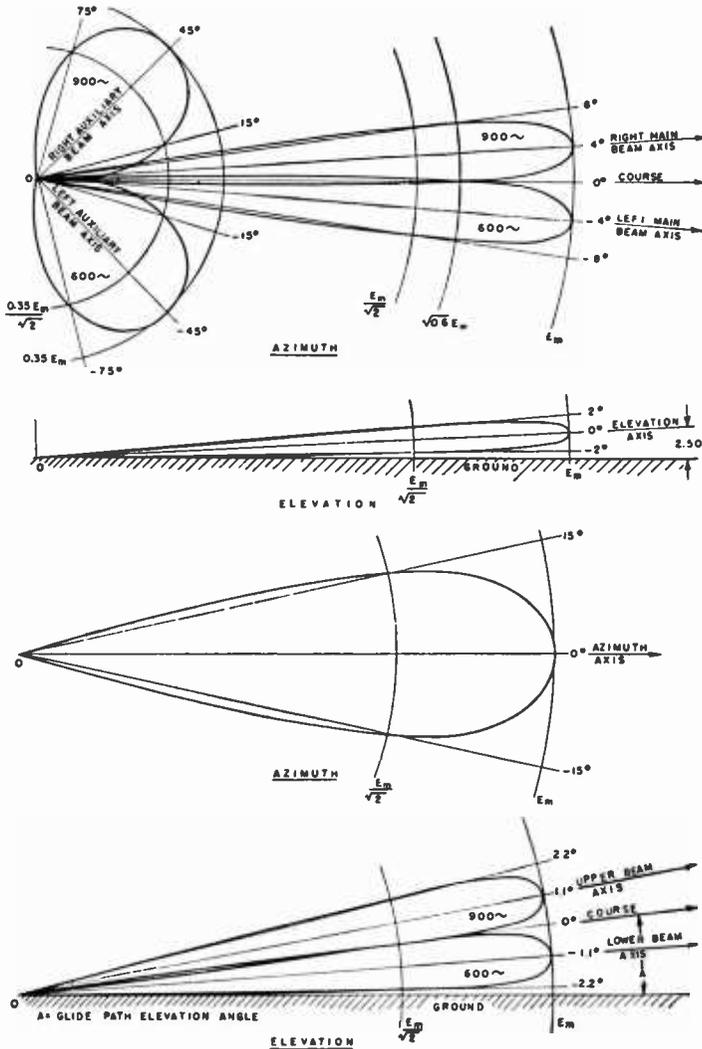


FIG. 34. Sperry microwave instrument landing system.

localizer operates on a frequency of 2,639, 2,639.5, or 2,640 Mc and the glide path on 2,616, 2,616.5 or 2,617 Mc. Each transmitter has a power output of 70 watts and is crystal-controlled at about 5 Mc, with frequency multipliers to the final carrier value. Modulation of the overlapping beams of each unit is at 600 and 900 cps, respectively, carried out by mechanical modulators. The glide-path antenna is a 12-ft vertical

parabola with two wave guides from the modulator feeding the two modulated waves into the mouth of the radiator near the focal point, but slightly off focus to form two overlapping beams. The localizer antenna consists of three main parts: a paraboloid of revolution (6 ft in diameter) which transmits the two main beams, and two smaller parabolas (6 ft high and 3 in. wide) which transmit two auxiliary side beams. The former provides the runway course; the latter, the wide azimuth coverage needed to enable the pilot to intercept the localizer beams without difficulty.

The operating principles of the system, as well as an appreciation of the precision provided, may be had from Fig. 34. Extensive flight tests have been in progress on this system for several years, at a large number of airports and under varying conditions of weather and of runway location. Straight, precise, runway localizer and landing paths are provided to a reliable distance range of 50 miles, under all weather conditions.

Other Systems. There have been many proposals for instrument landing systems, involving every means of guidance including nonradio methods. One system employed an air-borne direction finder and altimeter in combination with a system of ground position locator markers along the runway and its extension. Another utilized the magnetic field surrounding two horizontal multiconductor cables (laid on either side of the runway and its extension) to establish in space a path of constant electromagnetic field intensity (500 cps) which coincided with the desired landing path.

With the advances in radar during the war, it was to be expected that a large number of proposals would appear for the application of radar techniques to the aircraft landing problem. The more common combinations involve the use of an air-borne radar set in conjunction with man-made reflectors or transponder beacons outlining the airport approach and runways. A number of the proposals are very attractive for pilot use but begin to show deficiencies when applied to the automatic landing problem.

ABSOLUTE ALTIMETERS

48. F-m Altimeter. Altitude above ground is determined by measuring the difference in frequency between the transmitted wave and the ground-reflected wave for an f-m carrier. The equipment used comprises a transmitter, a receiver, and a frequency-measuring circuit operating a "terrain-clearance" meter.

In a typical commercial altimeter of this type (RCA) the transmitter operates on a power output of 0.1 watt and on a nominal carrier frequency of 440 Mc modulated at 120 cps. Two band widths are used (40 and 4 Mc), corresponding to two altitude ranges of 0 to 4,000 and 0 to 400 ft, respectively. The transmitter output is radiated downward by a $\lambda/2$ horizontal doublet mounted on one of the lower surfaces of the airplane, and the receiver is connected to a second $\lambda/2$ horizontal antenna similarly mounted and arranged for minimum coupling to the transmitting antenna. The direct and reflected waves picked up by the receiving antenna are fed to the detector, and the detector output is amplified and fed to a frequency-measuring circuit and meter indicator. It will be evident that the frequency of the signal output from the detector is equal to the instantaneous frequency difference existing between the direct and reflected waves and is directly proportional to the height above the ground. Thus the frequency meter may be calibrated directly in feet.

For any altitude, H , the frequency difference, f_d , is given by the expression

$$f_d = 2f_m B \times 10^6 \times \frac{2H}{C}$$

where C = velocity of light in the same system of units as H

f_m = modulator sweep frequency, cps

B = band width, Mc

Thus, for either altitude range, the frequency difference per foot of altitude is approximately 20 cycles. A high order of accuracy may thus be attained. The stated accuracies for the high- and low-altitude ranges are, respectively, ± 60 ft ± 5 per cent of the reading and ± 6 ft ± 5 per cent of the reading.

As the radiation toward the ground is directed over a rather large area, the indicated altitude above ground will be the average of the elevation covered by a cone of about 30 deg. Therefore buildings and other small obstacles affect the readings appreciably only at low altitudes. Clouds have negligible effect on the readings, since the reflection coefficient of radio waves from clouds is only about 0.01 per cent of that from land. The values of 60 and 6 ft represent the fixed errors of the equipment which are partly reduced when the aircraft is flying above reasonably rough terrain.

Many thousands of these altimeters were used both by the U.S. Armed Forces and British during the war, and a number have been produced for the commercial market.

49. Radar Pulse Altimeter. This altimeter, like the f-m type, utilizes the principle of emitted and reflected radio waves; however, the pulse techniques employed allow measurement to much higher altitude.

One radar altimeter now available commercially (RCA) operates at 440 Mc (10 watts peak power) and gives altitude readings from 0 to 40,000 ft with an accuracy of 0.25 per cent of reading ± 50 ft, but the error should not exceed 100 ft at 40,000 altitude. Altitude is shown on a cathode-ray screen directly in feet, and the reading is indicated by a "pip" which is caused by the echo signal and which appears at the proper location on a circular base line. When an aircraft is flying over mountainous terrain, the pulses sent out by the transmitter will be reflected from mountain peaks and from flatter areas, resulting in different reflection times and the consequent appearance of several different pips, thus indicating irregularities in the topography of the land.

The oscilloscope of the radar altimeter displays not only the returned echo but also the transmitted pulse, the latter serving as a reference indication. Since these two indications tend to merge at very low altitudes, a lower limit of altitude indication is imposed, approximately 25 ft.

50. Earlier Altimeters. Earlier absolute altimeter methods included the *capacitance altimeter* and the *sonic altimeter*. In the former, the distance from ground was measured by detecting the change in the electrical capacitance between two plates on the airplane as it approached the ground. In one form, this capacitance was made a part of a resonant circuit coupled to an extremely stable r-f oscillator, and a vacuum-tube voltmeter was used to record the voltage change as the approach to ground brought the resonant circuit into closer tune. In the sonic altimeter, the time taken by sound to reach the ground and return to the airplane was measured. In one model, two horns were used: one, driven by an electric trip relay and plunger, transmitted the sound wave; the other received the ground-reflected echo. A sound-delay filter permitted separation of the echo from the transmitted sound pulse down to 5 ft above the ground.

Neither of these two types appears to have present application. They were developed primarily for use in approach and landing and, in fact, are by nature restricted to low-altitude service. The ratio of weight and cost to service rendered appears to have ruled them out of serious consideration.

INTEGRATED SYSTEMS OF AIR NAVIGATION AND TRAFFIC CONTROL

51. Present CAA Traffic-control Procedures and Development Trends. The requirements for air traffic control, as envisaged by ICAO, were detailed in Sec. 2. The facilities utilized by the CAA in controlling air traffic along the U.S. airways and at airports were outlined in Sec. 2. A brief discussion of the traffic-control procedures currently followed will now be presented as the best means of demonstrating the complexities of the problem. Certain of the approaches for improving the efficiency of control will also be considered.

The most pressing need for air traffic control is when instrument flight conditions prevail.¹ Before taking off, the pilot of each aircraft intending to fly along the civil

¹ CAA refers to the regulations pertaining to instrument flight conditions as IFR and to those for contact flight conditions as CFR. See references to W. D. White and to the CAA bulletins on this subject at end of chapter.

airways under instrument flight conditions must file a flight plan showing the route to be followed and the desired cruising altitude. This flight plan is transmitted to the traffic-control center where it is posted on flight progress strips on a traffic-control board. The time of passing over each fix is estimated and, if there are no conflicts with other traffic, a clearance is issued. En route the pilot reports to the control center his time of passing over each fix allowing the control center to revise its estimates as necessary. Some of the rules employed in issuing clearances follow:

Two planes at the same altitude traveling in the same direction must be separated by 10-min flying time. Planes flying in opposite directions must be separated by 1,000 ft of altitude and are further separated by flying to the right of the on-course signal. Two planes taking off for different routes must be separated by at least 1 min while planes taking off for the same route at different altitudes must be separated by at least 3 min., if the first plane is going to the higher altitude. If the second plane is going to the higher altitude, this last rule must be modified to 5 min.

An important feature of the procedure is that, once a clearance is issued for a particular flight plan, the aircraft is protected so long as it adheres to this flight plan even though it subsequently loses all radio contact with the ground.

The most critical region of flight is in the approach zone to an airport. The problem is complicated by the fact that several different airways, as established by radio ranges and runway localizer beams, converge on the same airport. The standard procedure requires that the pilot communicate with the airport control tower when reaching the approach area and that he maintain a "holding" procedure under the instructions of the control tower until cleared for landing. When cleared by the tower, the pilot is required to start the approach glide from an altitude and position such that he be at an altitude of about 1,000 ft and in a glide of about 2.5 deg just beyond the marker of the ILS.

Immediately preceding the final approach, the aircraft is required to fly along an outward course for a distance of 10 or 12 miles from the airport, thus enabling the pilot to orient himself and to prepare for a smooth approach. The pilot then executes a 180-deg procedural turn and returns on the final approach path. A number of time-consuming delays are introduced in carrying out this type of instrument approach. No other aircraft can be cleared below the 3,000-ft level until the landing aircraft is actually on the ground. In addition, since aircraft awaiting their turn to land are required to fly a holding course having a 4-min lap, an aircraft, when cleared for landing, must necessarily complete this course in order to enter the final approach path at the right point and in the proper direction. The time which will elapse between the instant of clearance and the start of the final approach is thus a variable factor.

A special procedure utilizing the air-borne ADF, operating in conjunction with two ground compass locator stations, has been designed to reduce the time delays inherent in the standard approach system, and thus expedites aircraft landings. This improved procedure is in experimental operation at La Guardia Field, New York. Aircraft awaiting clearance for landing fly a "race-track" holding course similar to the one employed with the standard system, but defined by the compass locator stations. The end of this course nearest the airport is adjacent to the landing strip instead of being nearly 15 miles distant, as is the case with the standard system. The axis of the holding course crosses the runway center line at an angle. With this arrangement, aircraft flying the holding course home alternately on the two compass locator stations until such time as the ground controller selects a particular aircraft for a landing. The pilot then reports when he passes over the outer compass station and is assigned such a heading that the execution of a 180-deg right turn at a definite later time will bring the aircraft directly in line for the final approach. Figure 35 shows that the size of this final turning loop can be controlled by varying the outbound heading and the time of the right turn to allow for any random departure from the holding course at the time the aircraft is first ordered to report. Although this system allows a theoretical headway of 3 min on the approach path, take-offs, which are necessarily accomplished

between landings, prevent this headway from being achieved in actual practice. For this reason, the actual minimum time interval between successive final approaches is never less than 4 min.

52. Present-day Problems. With the rapidly increasing congestion at airports, such that the peak rates of landings and takeoffs at a busy airport may reach 100 (total) per hour within the next few years, present control methods are seen to be inadequate. New techniques under proposal include: (1) a pictorial display to the pilot (by radar or television) of the position of other aircraft in his vicinity; (2) division of the airways on a block-system basis with a light system in the aircraft, operated from the ground, clearing the pilot for further flight or advising a holding procedure or an alternate routing; and (3) time synchronization of each aircraft's position over strategic control points. The service values of these proposals are yet to be tested, and it may be expected that a combination may evolve, including also some form of collision-prevention aid within the aircraft itself and operating independently of any ground control.

A third phase of air traffic control, apparently insufficiently considered until quite recently, is the adequate handling of the ground movement of aircraft and service

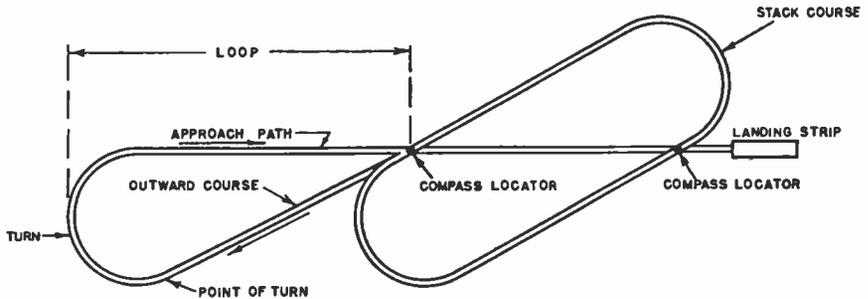


Fig. 35. Improved procedure for expediting aircraft landings. (Courtesy of W. D. White.)

vehicles at the airport under poor visibility conditions. It is now generally agreed that there are two major requirements to be met in establishing a satisfactory control over ground movements: (1) the determination of obstructions on the runways in use and (2) a survey of the complete movement area.

By comparison with the extensive automatic systems now needed for the proper control of ground traffic at airports, the facilities at present available for that purpose are extremely meager. They consist solely of instructions to the pilot from the control tower, by means of radiotelephone or light signals. The control-tower operator is apprised of the position of the aircraft either by visual observation or by radiotelephone communication with the pilot. The difficulty of clearing the runway of landed aircraft constitutes the greatest drawback to the present system and is greatly accentuated whenever poor visibility prevents direct observation by the control-tower operator. Take-off headway is coordinated between the control-tower operator and the pilots of aircraft in flight. The latter report by radiotelephone to the control-tower operator who supplements this verbal information with visual observation of the airfield and with reports from the pilot of the departing aircraft. This procedure requires that a large amount of air space be held clear at all times, thus slowing the movement of traffic; it is open to the further objection that errors may be made by the reporting aircraft pilots.

The problem is one of surveillance, identification, and control. Inasmuch as the ground positions of aircraft and vehicles at any active airport are changing from minute to minute, any satisfactory system of surveillance must present the controller with a practically continuous flow of information. Several different methods have been considered for providing the necessary information regarding the movement of ground

traffic: (1) the use of personnel posted at strategic points, (2) traffic pads that operate electrical indicators at the control point, (3) optical beams, including infrared, across runways to actuate photocells, (4) inductive pickup of ignition fields, (5) microphonic pickup of engine noise, (6) simple radar scanning of runways in use, and (7) radar scanning of the whole movement area and presentation on plan-position-indicator (PPI) oscilloscopes. The last two methods appear to have the best chance for ultimate success.

It is essential that the controller not only distinguish between certain fixed objects and the aircraft and vehicles under control but he must be able to identify positively each aircraft and vehicle. No satisfactory solution of how to accomplish this on the position indicator (with the two radar methods) is in sight. Ground direction finding, radar using special responders installed in all vehicles, and coded corner reflectors, have all been considered, but none of these methods is at present capable of the necessary resolution.

The issuance of instructions for the movement of ground traffic also poses a considerable problem. Possible methods include the use of (1) "follow-me" vehicles, (2) pyrotechnics, (3) radiotelephone from the control point to aircraft or vehicles, (4) leader cables, (5) stop-and-go traffic lights, and (6) taxi-track lights. Lighting systems that are switched from the control point appear to offer the most likely means of controlling aircraft. Although the illumination produced by such systems may not be visible under all conditions, it is probable that a lighting system can be developed which is satisfactory for most weather conditions. Photocell operation has been suggested as a possible adjunct to such a system.

53. Radar Aids to Air Navigation and Control. Descriptions of special applications of radar to the needs of civil aviation and general comments as to potential applications have been presented in the preceding paragraphs. In considering potential applications, particularly with respect to air navigation and control, it is essential to distinguish between pure radar systems and those involving interrogators and transponders.¹

In pure radar systems, the search or surveillance functions (whether by air-borne or ground equipment) are carried out by reflection of the transmitted wave from natural targets. For air-borne equipment, navigation is by landscape (as presented on a cathode-ray indicator) which may be augmented by man-made reflectors such as trihedral and biconical corner reflectors. For ground equipment, the airplane serves as the reflector with ground objects providing the natural fixes. In general, the antenna directivity must be high to yield accurate directional information, the p.r.f. must be high to afford indicator persistence, the pulse length must be short for good range resolution, and the peak power must be high to give adequate range of operation. Because of the usual requirement for small antenna structures, the trend has been toward the use of the highest possible frequencies, even though these generally exhibit more variable propagation characteristics in varying weather.

54. Interrogation and Transponder Systems. The design requirements in this case may be quite different. For both the air-borne and the ground equipment, it is generally preferable to operate the interrogator and the beacon transponder on different frequencies, thereby avoiding clutter from natural reflectors which may operate to obscure the desired indications. Since the transmission is one-way on each half of the round trip of a signal, field intensity varies inversely as the distance rather than inversely as the square of the distance for pure radar systems. Thus, transmitter peak-power requirements are considerably lower than for pure radar systems; in fact, a definite balance must exist between the power of the transmitters and the sensitivity of the receivers in the entire interrogator-responder network, or a particular transponder beacon may transmit signals which serve no purpose but to produce interference with other units in the system. Pulse lengths are in general greater and, at any rate, must appear in coded groups to provide identification of the various transponders in the system. The p.r.f. must be low so that a single beacon may be interrogated by

¹ See reference to F. A. Darwin at end of chapter.

a large number of interrogators and, furthermore, so that the receiver indicator will not be saturated by the responder signals at a busy location along the system network. Since a great deal depends upon the maintenance of balance in the system, the choice of operating frequency must be such as to guarantee stable propagation characteristics.

It is evident, then, that any combination of the two systems must deal carefully with a number of very important engineering compromises. The net result is that it may prove desirable to employ dual air-borne or ground equipments, embodying both the pure radar and interrogator-responder types, in applying radar to carry out its potential air navigation and control functions.

Ground Surveillance Radar. The search system in the GCA equipment (see Sec. 45) serves as an excellent example of a ground-station surveillance radar.

Air-borne Radar Search with Interrogator-Responder Adjunct. A wartime air-borne radar set which is currently undergoing redevelopment for miniaturization and application to civil air transport use is the APS-10. Weighing 125 lb in its original form, one commercial development (Allison Radar & Equipment Company) indicates the possibility of a 65-lb unit which can be housed in a container 19 in. wide and about 30 in. vertical depth (including its twin 18-in. reflectors). The entire assembly may thus be mounted in the nose of any four-engine transport. The set operates at a frequency of 9,375 Mc, and requires only 600 watts of input power (400 cps, 115 volts, single-phase).

The following description applies to the wartime APS-10 set:

The APS-10 transmitter has a peak-power output of 4 to 8 kw, a pulse width of 0.84 μ sec, and a p.r.f. of 810 per second during search operation. The pulse width is increased and pulse-repetition frequency decreased for beacon interrogation. The receiver is a super-heterodyne with local oscillator and crystal mixer. The i-f band width is 6 Mc (30 Mc center frequency). An a-f-c system controls the local oscillator repeller voltage (during search operation) so as to maintain the correct i-f frequency by correcting for frequency drift of the magnetron transmitter and of the local oscillator. This control is effective in holding the i-f input at 30 Mc to within 0.5 Mc. A separate local oscillator is used during beacon interrogation so that reception is on a separate frequency; a.f.c. cannot be used during such operation.

The APS-10 indicator contains the viewing oscilloscope, a 5-in. magnetic deflection cathode-ray tube with conventional PPI display. In addition to the tube it also contains a final video stage for the received signal, deflection coils for the scope, and controls for properly centering, focusing, and otherwise adjusting the radar pattern for maximum clarity. Range measurement is by electronic range marks which are passed through the video amplifier and appear as concentric circles on the scope face with the signals. Azimuth measurement is taken from radial lines etched on an amber filter placed on the front of the indicator.

The APS-10 synchronizer contains all the timing and control circuits for operation of the equipment. It also mounts the majority of operating controls such as range selector, video gain, antenna turn and tilt, and a switch for shifting from search to beacon operation. The sweep circuits of the indicator are developed in the synchronizer and fed to the rotor of a synchro in the antenna. The synchronizer also contains a meter for indicating the vertical tilt of the antenna.

The APS-10 antenna is a 3-cm dipole backed by a 20-in. parabolic reflector. It is equipped with two 24-volt d-c motors for rotation in azimuth and elevation. Normal speed of azimuth rotation is 30 rpm and the tilt may be adjusted from +1.5 to about 30 deg. Rotation in the horizontal plane is accomplished by a rotating joint consisting of a split section of circular wave guide, and in the vertical plane by a nodding joint. The parabola has been modified by the addition of a horizontal metallic strip, known as a "spoiler," to broaden the normally narrow vertical beam width to a cosecant-squared pattern. In this way, uniform received signal strength is obtained from ground objects over a distance of several miles in the direction in which the antenna is pointed. The antenna beam width in the horizontal plane is held to 8 deg so that sharp PPI pictures are produced. The antenna also contains a selsyn, the rotor of which turns with the azimuth drive motor. Since the rotor coil carries the sweep circuit currents from the synchronizer, varying currents are induced in the stator coils dependent upon the position of the antenna. In parallel with these stator windings is the three-winding deflection coil of the indicator scope. Thus the effect is to rotate the PPI radial sweep in phase with the antenna.

In the Allison modification of the APS-10 set, the two antenna dishes mounted back to back are set at slightly different angles so that two separate fields of search are

covered, one beam sweeping directly ahead and up, the other (tilted downward) sweeping the terrain. Each dish sweeps through a 180-deg forward arc in azimuth, at 30 scans per minute for each dish. The width of each beam in elevation is approximately 6 deg. An r-f switch cuts in power to each dish as it enters its 180-deg azimuth sweep. An important feature of the set is the stabilization of the antenna assembly by a single gyroscope installed between them. The scope is an RCA "A" type of high persistence and high intensity. Results of the search are switched alternately to this indicator, objects picked up by the upper beam appearing with four times the intensity of those from the terrain-scanning beam. The range of the ground scan is limited to 5,000 ft from the plane. Apparently, this set does not provide separate frequency reception during beacon operation.

In tests of the wartime APS-10 by the CAA, reflections from aircraft have ranged from 1 to 5 miles depending upon size and reflections from ground objects have been

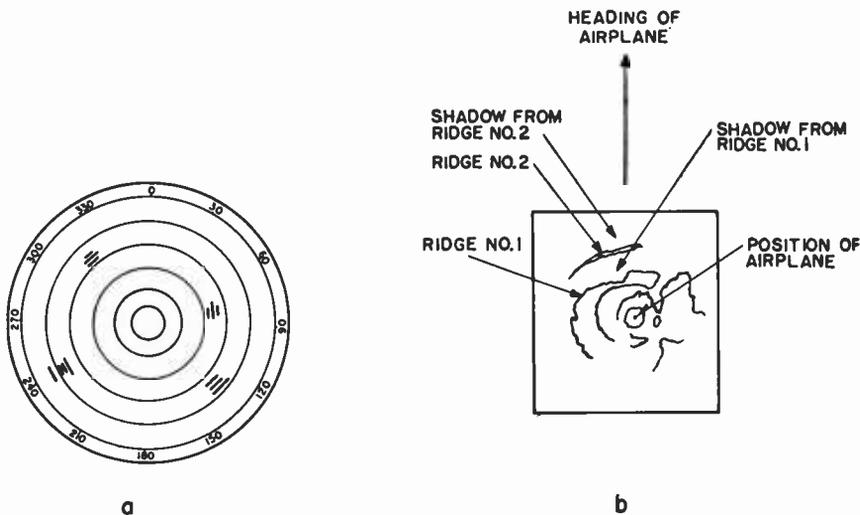


FIG. 36. Typical presentations on the wartime APS-10 viewing screen. (a) Beacon replies; (b) expanded view of search picture.

received up to 50 miles in range. Ground beacons would have an equivalent range of 100 miles. Storm clouds were detected at ranges of some 20 miles. Typical presentations on the PPI screen are shown in Fig. 36. Here (a) shows the characteristic coded responses of ground beacon stations where range is measured (radially) to the beginning of the inner pulse, and (b) shows the enlarged central position of the screen during search operation.

The utility of an air-borne radar set of the type described is multifold. As a search radar, it constitutes a self-contained aid, providing position data of value in the more accurate reporting of arrival over airway traffic-control points. It yields information on terrain clearance and provides "present-position" data regarding neighboring airplanes, thereby reducing danger of collision. It also yields limited information on approaching storms. In conjunction with ground beacon stations at important points along the airways or in the vicinity of airports, it would greatly reduce the hazards arising from falsely assumed position or from faulty flying procedure in airport approach zones.

Proposals have been made to outline all airways, airports, and even landing strips with low-powered radar beacons, so that an airplane equipped with radar could navigate as by lights from beacon to beacon. Even with the present facilities along the

airways, such a proposal seems to involve unnecessary complications. With the advent of the omnirange and DME, much of the purpose of such a system would vanish.

55. Navar System. The Federal Telecommunication Laboratories, Inc., have proposed and are developing a series of flexible coordinated radio aids to air navigation and traffic control of which the R/θ navigational system forms an essential core. The key facilities which are to comprise the system and which have been demonstrated experimentally, are the F. T. & R. omnirange (Sec. 24), DME (Secs. 5 and 38), and ground surveillance radar (Sec. 53) used in conjunction with responding beacons in the aircraft. Each facility of the system, when perfected, is seen as fitting into a carefully integrated plan directed to meeting all the foreseeable requirements of air-transport operation.

The air-borne radio sets required by this system and the various services obtained are indicated in Fig. 37. The steps anticipated in progressively more comprehensive

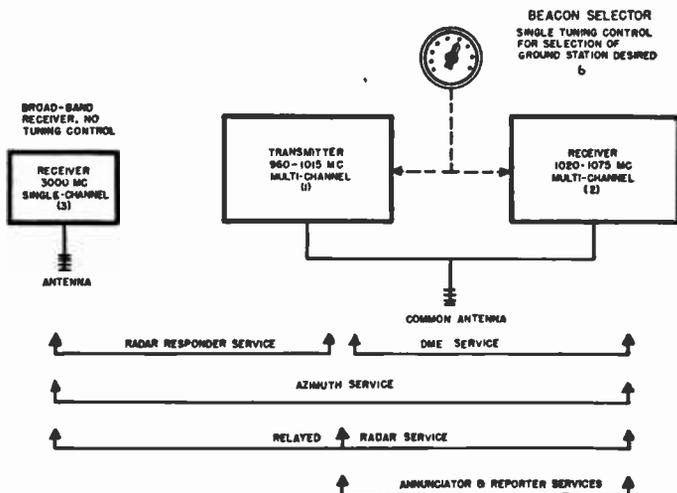


FIG. 37. Features of Navar air-borne radio sets. (Courtesy of Federal Telecommunication Laboratories.)

adoption of the system are shown at (a), (b), and (c) of Fig. 38. The transmitter (1) and receiver (2) constitute the normal DME equipment required in air-borne use of the CAA facilities and may be used, per se, in conjunction with the CAA omnirange azimuth meter. The additional airborne equipment required by the Navar system is the 3,000-Mc receiver (3) and the indicators and displays to be described. The airplane requires no bulky rotating directional antennas with this system.

Phase a. Addition of Azimuth Meter Service. This was described in Sec. 24. The ground surveillance radar transmits directional signals which, operating in conjunction with synchronizing north signals from the ground DME beacon, provide azimuth information. Upon the reception of each radar challenge pulse on the aircraft receiver (3), a time-modulator unit, which normally creates the DME interrogation signals continually, produces a coded response signal for triggering transmitter 1 to produce artificial radar responses used in the display on the ground. The radiation from transmitter 2 may bear a characteristic code for identification and a second coded signal (operated from the airplane's barometric altimeter) to yield altitude information on the ground. The ground surveillance radar receives the replies from the aircraft responders and displays these replies as echoes on a plan position oscilloscope. By using a multiple-channel receiver on the ground, it is possible to have either a "selec-

tive" beacon radar display, which shows only airplanes responding on a given channel, or a "comprehensive" display, which shows all airplanes regardless of their radio channel selection, or both displays. The "azimuth meter" service may be regarded as a by-product, available inexpensively, where ground radar surveillance sets are installed. The air-borne azimuth meter is made to receive only the "north" and directional signals corresponding to the selected ground surveillance radar station, as follows:

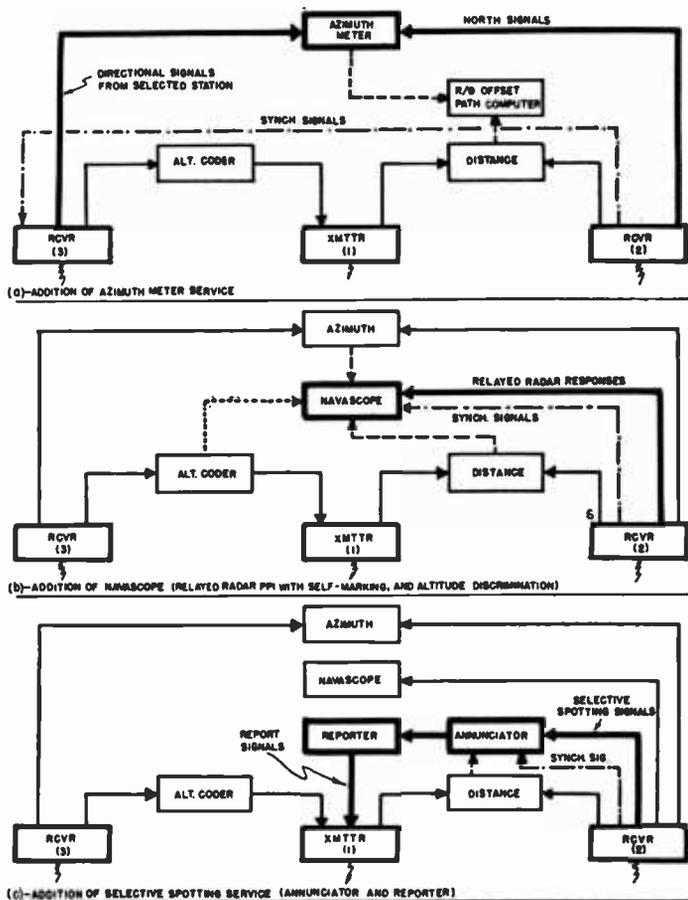


FIG. 38. Progressive consolidation of Navar air-borne equipment. (Courtesy of Federal Telecommunication Laboratories.)

Receiver 2 picks up the north signal only from the desired station but also receives the DME transponder beacon signals; to separate the former from the latter, selective spacing of the north signals is employed (a short train of 40 to 50 pulses uniformly spaced 20 μ sec apart) with a tuned (50-kc) control circuit in the output of receiver 2 building up sufficient voltage after about 20 of these pulses to send a north signal to the azimuth meter. Receiver 3 picks up directional signals from any ground station within its range being single-channel broad-band. The particular signals from the selected station are chosen by means of a second synchronizing signal from receiver 2

and derived by a coincidence gate from the north synchronizing signal. The azimuth meter receives only these directional signals.

Phase b. Addition of Relayed Radar PPI on Airplanes (Navascope). A pictorial or radar-type presentation (PPI) is provided for each airplane, showing locations of other airplanes as well as the location of the pilot's own airplane, as seen at the ground radar station. All the replies received at the ground for the "comprehensive" display are rebroadcast aloft on the DME beacon transmitter and received on the airplane on receiver 2. To duplicate in the aircraft the ground PPI display, the reference and control signals needed for synchronization (for time base reference and rotational deflection) are already available (for operation of the azimuth meter). Spots of light appearing on the PPI display will represent all aircraft equipped with responder equipment; the particular spot corresponding to the pilot's own airplane may be distinguished by extra size and brilliance on the basis of control signals from the independent self-distance and self-azimuth meters.

Since only those airplanes at the approximate altitude of the pilot's own airplane are immediate sources of danger, the pilot may use his own altitude coder to control a PPI altitude discriminator. This is possible because all the altitude-coded replies from the different airplanes, although transmitted to ground on different radio channels, are retransmitted aloft on all channels. In passing from one altitude to another, the pilot can selectively investigate other vertical layers.

Phase c. Addition of Selective Spotting Services. Here, selective spotting facilities are incorporated, whereby a sort of radio "finger" can be put on any selected airplane appearing on the ground station PPI. The ground controller adjusts an azimuth knob which swings a radial beam of light to cross the selected spot of light on the PPI. This automatically causes a highly directional antenna of a special ground transmitter (on the proper DME channel) to point accurately in the direction of the selected airplane. Similarly, adjustment of a distance knob to cause a circular beam of light also to cross the selected spot of light on the PPI causes the special transmitter to send out signals which are timed (with reference to the existing synchronizing signals) to correspond to the actual distance of the selected airplane. The air-borne self-distance meter and the synchronizing signals are used to operate a distance discriminator which allows only the selected airplane to receive the spotting signals.

In this way coded traffic dispatching signals may be produced on an annunciator in the selected airplane, or the airplane's transmitter may be caused to emit a brief reply as to identification and altitude. The value of this method of selective signaling in air traffic control is self-evident. Either of the methods of ground control mentioned in Sec. 51, *i.e.*, time synchronization or block signaling, will require selective signaling from the ground.

56. Sperry Coordinated Air-traffic-control System. With its microwave instrument landing system as the foundation, the Sperry Gyroscope Company has proposed and is developing an over-all automatic air-traffic-control system based on microwaves. After a careful study of the factors involved, Sperry has adopted the following design goals:

1. All aircraft must know their own azimuth position and distance in relation to the traffic-control center. The airplane should receive this information automatically and in such a manner as to be easily seen. The use of a visual presentation in polar coordinate form seems most desirable. Azimuth and distance information in the aircraft must be in a suitable form so that automatic flight paths can be achieved by operation of the automatic pilot from the ground control signals.

2. The ground control station (the center of the polar coordinate system) must have as a minimum the following information regarding all aircraft within the controlled area: (a) azimuth (omnirange), (b) distance, (c) altitude, and (d) identification.

3. The information at the ground control station should be presented automatically so that a pictorial presentation of the traffic in the controlled area can be continually observed. This infers that any number of aircraft can be controlled and that the system must not have a saturation level because of technical shortcomings.

4. The coordinated system must include instrument landing equipment which will furnish signals accurate enough to guide the airplane all the way to its stopping point on the runway. These signals must be suitable for use with a precision automatic pilot so that eventually completely automatic approach and landing can be accomplished.

5. The aircraft must be able to communicate with the ground, if necessary, but all routine communication should be automatic and voice communication should be employed only when emergencies arise.

6. The basic elements of the system (distance indication, omniazimuth, and instrument landing) should be so integrated that the minimum of radio spectrum is used, and the minimum of air-borne equipment is required. It is necessary to have one multipurpose transmitter and one multipurpose receiver in the aircraft for several simultaneous operations. Practically, this requires that all traffic-control data transmission and voice-communication channels be in the same part of the radio spectrum.

7. The system should be flexible enough to permit all aircraft, even those not under its immediate control, to utilize desired portions of the total information. Smaller aircraft with limited equipment should be considered in the over-all control plan.

8. Finally, the system should possess possibilities for continued expansion in every respect.

The *azimuth-determining system* was described in Sec. 24. A nondirectional reference-phase signal and the variable phase of a rotating field signal, produced by mechanical rotation of an antenna, are compared on the aircraft (as in the CAA omnirange). The Sperry system eliminates the phasing difficulties encountered when producing rotating field patterns with stationary antenna arrays and obviates the necessity for adjustments subsequent to the initial installation of the antenna. The entire equipment is sufficiently compact and light to permit its installation in a weatherproof housing on poles, masts, tall buildings, etc. Because the radiation is directed away from the ground, siting problems are minimized.

The *distance-measuring system* was described in Sec. 38. It is planned to use the nondirectional carrier of the omnidirectional range for retransmitting the signals received from the aircraft. Saturation of this channel is not anticipated, since the aircraft transmitters will in general have different modulating frequencies. A major advantage for this method of distance measurement is that the indications are also available at the ground station.

Azimuth and distance scanning are possible with this system by an ingenious arrangement that affords complete position information on the ground corresponding to all aircraft within the area. A separate microwave link is required for this service. The ground transmitter of this link is modulated by a signal which is derived from the reference signal of the omnirange but is caused to vary continuously in phase at a constant rate of, say, 60 deg per sec by passing the omnirange reference signal through a continuously varying phase shifter. This signal will be referred to as the variable-phase link signal. Ground-station transmission is from a nondirectional antenna. On the aircraft, there is already available the variable-phase signal from the omnirange, the phase of which is a direct function of the bearing of the aircraft. A second phase-comparison circuit is provided on the aircraft which compares the variable-phase signal from the omnirange receiver with the variable-phase link signal.

The aircraft transmitter for the separate microwave link is normally in "stand-by" condition. Its modulation consists of a frequency corresponding to the distance from the ground station and developed in the aircraft as a result of the distance-measuring equipment. As the variable-phase link signal from the ground station passes through phase synchronism with the variable-phase omnirange signal (at the aircraft), the output of the second phase-comparison circuit is made to trigger the transmitter. The ground-station receiver then obtains a signal which indicates the distance of the aircraft.

The arrangement described is seen to comprise an azimuth-scanning system.¹ Since the signaling phase shifter is varied at a constant rate, the transmissions from aircraft

¹ It should be noted that this method of azimuth scanning is applicable to any phase-comparison type omnirange, such as the v-h-f omnirange of the CAA.

at different azimuth bearings will be separated from each other on a time sequence basis. The distances of the aircraft will be distinguished from each other on a modulating frequency basis. If two or more aircraft appear at the same azimuth bearing, their transmitters will respond at the same time but will show different modulations, according to their distance.

An added feature of the azimuth-scanning system is that the distance-data modulations are retransmitted on the ground-station transmitter of the microwave link on a time sequence corresponding to the azimuth bearings of the several aircraft transmissions received as a result of the azimuth scan. The aircraft receiver thus has information as to the bearing and distance (with respect to the ground station) of all airplanes within the service area of the ground station. It is easy to select from these data only those signals representing aircraft within a given radius of the pilot's own aircraft. Azimuth information for this area can be filtered out by observing signals just prior and just after the interrogation of the pilot's own transmitter by the ground scan. Distance filtering is accomplished by a filter passing modulation frequencies in a small band centering on the distance modulation for the pilot's own aircraft. Thus, the pilot of each aircraft may select a display for his own area (which moves with him) and may place his own craft in the center or in any other position in the display.

Ground-to-aircraft signaling, on a selective basis, is feasible with this system through the use of a second relay on the aircraft operated from a circuit in the link receiver output which compares the aircraft's own distance modulation for equivalence with the distance modulation received from the ground station of the link. This relay may be connected in series with the relay operated by the azimuth scan. When the operation of the two relays coincide, an annunciator may be actuated by a coded sequence from the ground station. It is evident that the ground-station controller may thus send simple traffic-control dispatches to any selected airplane by sending the coded intelligence signals on the correct distance modulation and at the correct instant of the azimuth scan.

As indicated in connection with the discussion of selective signaling in the Navar system, this feature is essential in any system of ground control of air traffic. The intelligence transmitted to the selected aircraft may be warning signals based on the future position of other aircraft in the area as determined from automatic computer information at the ground station. Such warning signals may include instructions as to evasive action necessary to avoid collision. Furthermore, the intelligence may comprise actual control signal to the automatic pilot of the selected airplane so that the aircraft may be guided automatically according to computer information on the ground. Sperry has been experimenting with various automatic means for feeding the signaling intelligence, on a selective basis, into the system, so as to control air traffic automatically.

57. Hazeltine Lanac System. In the Lanac system the basic principle of operation is that of transmitting a challenging signal from aircraft and thereby actuating automatic equipment on other aircraft or on the ground, which in turn transmits a reply. The system thus provides for the automatic exchange of data between various aircraft, and between aircraft and the ground. In essence, it provides immediate information concerning the presence of other aircraft or ground stations, their exact locations, and their complete identification. The employment of coding and decoding techniques permits the exchange of information in a channelized and highly concentrated manner, thus eliminating spurious and unwanted replies. Based on the advantage of a triggered-beacon system, in that the responding signal can be made to carry information originating at the target, proponents of the system believe that Lanac can serve as an effective adjunct to the other facilities and by ground controllers in their control of air traffic and, simultaneously, furnish anticollision service. Obstacle-warning beacons installed on mountain tops or tall buildings constitute the only ground facilities required for this anticollision function.

The air-borne receiving equipment displays the signals from a replier and also, without any manual adjustments, actuates an indicating mechanism which gives a direct

reading of the distance from the challenging transmitter. The bearing of the replier can be determined by means of the challenger's antenna-bearing indicator. Provision is made in the air-borne replier for transmitting (on the challenger) a special identity code for special control purposes. This code, which utilizes double pulses, is switched into operation only on instructions from the ground controller. In addition, there is a special emergency control in the air-borne equipment which is to be used only when an aircraft is in distress. When this control is operated, the air-borne replier transmits a distinctive pulse group which notifies both air-borne and ground challengers that the aircraft is in trouble.

For air traffic control, the same air-borne replier used in conjunction with air-borne challengers is operated in conjunction with two ground challengers and a common control unit. One of these ground challengers displays replies on several PPI oscilloscopes for general detection, location, and surveillance, each oscilloscope displaying the signals corresponding to a particular vertical layer. The other ground challenger uses

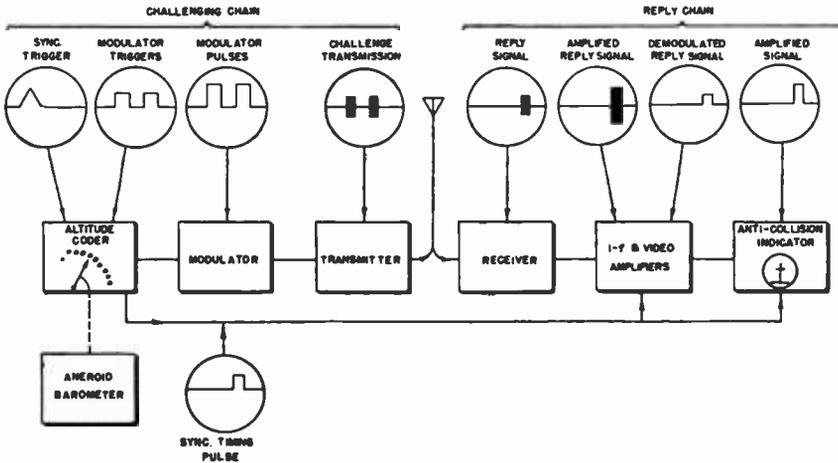


FIG. 39. Block diagram of air-borne Lanac challenger. (Courtesy of Hazeltine Electronics Corporation.)

an A scope, which gives range indications along a horizontal base line, for the purpose of identifying specific aircraft. The antenna of the latter challenger is manually rotatable and permits the operator to select any desired azimuth and lamina and thus to obtain the display and identification of the coded pulses from a particular aircraft.

The block diagrams of Figs. 39 and 40 show the basic elements of the Lanac system. Figure 39 shows the transmitting and receiving equipment comprising the challenger's installation, together with the pulse forms employed in different portions of the circuit. Figure 40 gives similar information concerning the replier's installation. (1) A train of pulse pairs is generated by the challenger's coder unit, the spacing characteristics of these pulses being determined by the coding to be used for the particular information desired. The operation of the coder is controlled by the aneroid barometer which thus determines the coding in accordance with the requirements of the different altitude layers. The coded pulse pairs are used to trigger the modulator of the challenger and to control the base-line sweep of the oscilloscope indicator on which the replies are to be displayed. (2) The pulse pairs produced by the coder cause the modulator to generate 1- μ sec pulses which modulate the transmitter oscillator and cause the latter to generate high-power r-f pulse pairs for radiation from the antenna. (3) All repliers within line-of-sight range of the challenger receive the challenging pulse pairs by means of nondirectional antenna system, an r-f system, and a wide-band i-f amplifier. The received pulse pairs are then demodulated by a diode detector and fed to pulse-code discriminator stages. As automatically determined

by the type of information desired, the discriminator or decoder circuits reject all pulse pairs except those having width and spacing characteristics acceptable to the particular replier. (4) Properly coded pulse pairs are accepted by a corresponding decoder circuit which permits them to trigger a reply-pulse generator. Each time the latter is triggered, it generates one or more pulses, the number, width, and spacing characteristics of which are dictated by the type of information to be supplied. For example, each air-borne replier is capable of transmitting a special signal whenever an aircraft is in distress. (5) For the purpose of providing exact identification, the train of reply pulses is varied in one char-

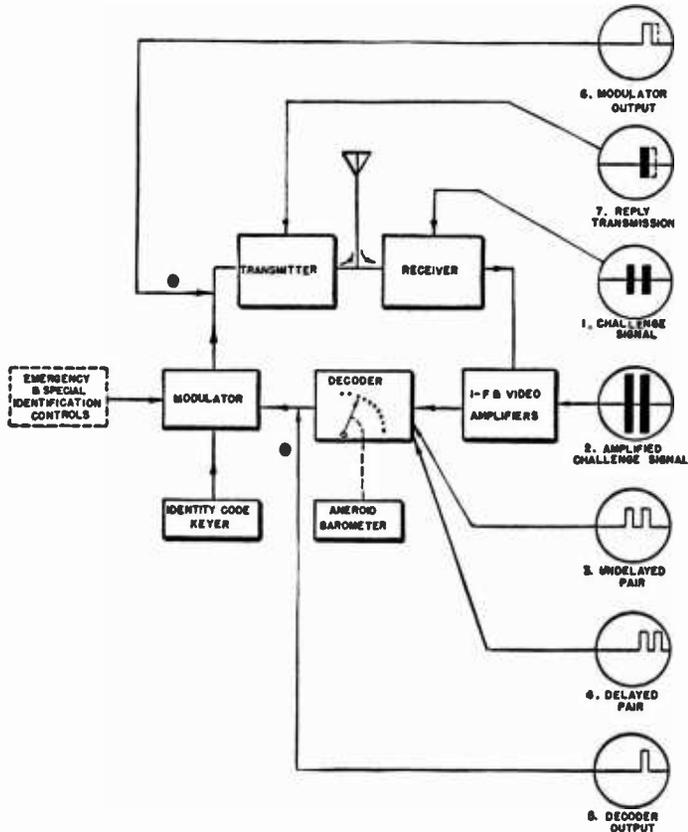


FIG. 40. Block diagram of air-borne Lanac replier. (Courtesy of Hazeltine Electronics Corporation.)

acteristic, for a visually perceptible fraction of a second, corresponding to the key-down periods of the international telegraph code. In this manner, the reply-pulse train, although uninterrupted, carries an easily read three-letter-code sequence which is repeated continuously as long as the challenge and reply operation continues. (6) The output of the reply-pulse generator is applied to the modulator which in turn causes the replier's transmitter oscillator to produce an r-f pulse at a different r.f. from that employed by the challenger. This pulse is radiated from a nondirectional antenna. (7) The challenger's receiver, which is closely similar to that of the replier, receives the reply-pulse signals and suitably amplifies the pulses for display on an associated indicator.

58. RCA Teleran System. Teleran (Television-Radar Air Navigation) is a system experimentally demonstrated by RCA for providing a visual presentation to the pilot

of the position of his airplane in relation to adjacent aircraft and terrain. Essentially, the equipment consists of an air-borne radar transponder and a television receiver. The ground equipment is basically a surveillance radar set and a television transmitter.

Signals from the surveillance radar transmitter are reflected by all aircraft within the area, the transponder on any particular airplane injecting a signal identifying that plane. When the signal is reflected to the radar receiver on the ground, the signal is presented on the screen of the usual cathode-ray tube. A television camera views this screen, and the television transmitter broadcasts the image upon which a map has been superposed. The image is received and displayed by the television receiver for the information of the aircrew. All aircraft in the area are presented as white spots on the screen; each craft is identified on its own scope by a radial line which passes through its pip. Although use of a persistent phosphor on the screen will cause each pip to show a trail that gives an indication of heading, it is also proposed to place before the viewing screen a transparent disk, ruled with parallel lines and positioned through a servo link with the stabilized gyro compass, to assure prompt indication of changes of heading.

To simplify the presentation, the Teleran system proposes to separate the radar echoes according to altitude and to transmit a separate picture for each of the following altitude levels: 0 to 2,000 ft, 2,000 to 4,000 ft, 4,000 to 6,000 ft, 6,000 to 8,000 ft, 8,000 to 10,000 ft, 10,000 to 15,000 ft, 15,000 to 20,000 ft, 20,000 to 30,000 ft.

At least 500 ft of overlap would be allowed on each side of the nominal vertical layer boundary to prevent an abrupt picture change in ascending or descending aircraft; e.g., an airplane at 2,400 ft would appear in both the 0 to 2,000- and 2,000 to 4,000-ft pictures.

As a conservative estimate, proponents of the system have set a 50-mile radius as the service area for each transmitter, with additional transmitters employed where natural obstacles would cause reduced range, particularly in regard to low-flying craft.

Use of the system has been proposed for en-route navigation, airways traffic control, airport approach and local traffic control, instrument approach and landings, and taxing control. It is also proposed for automatic traffic control, the transmission of weather data, etc.

In en-route navigation, conducted largely at high altitudes, the airways and their headings are displayed, together with the frequency channels to be used, the wind data, barometric correction, and the positions and headings of the pilot's own airplane and of adjacent aircraft. Topographical data, being of comparatively little interest, might be omitted.

Azimuth information is obtained by transmitting an omnidirectional pulse from the television transmitter at the instant the radar antenna is pointing north, and making an electronic measurement in the plane to indicate the time interval between the north pulse on the television channel and the direct interrogation pulse on the radar beam. Distance information is obtained by using the airplane's transponder as an interrogator after it replies to the ground station, the interrogating pulses being reradiated from the ground via the television channel.

Airways traffic control is handled by obtaining the location of all aircraft in the area with the surveillance radar, including craft without radar transponders. The coordinates of each aircraft being known, the progress of each is checked visually by the aircrew and ground personnel to avoid collisions. Where density of traffic requires tight control, the individual aircraft may be tracked and the data fed to a computer. As the craft progresses, it may be considered as moving through successive fixed blocks of space or as surrounded by a moving block of space, either concept being suited to the Teleran presentation.

Weather data can be supplied to the aircrew in the form of weather maps for different altitudes, presented at specific times on a separate channel.

For airport approach and local traffic control, a 20-mile radius is proposed. Some indication of terrain is planned, as well as a visual showing of approach paths.

For instrument approach and landings, it is proposed to show the extended runway

as a vertical line on the cathode-ray tube, mileage markings giving the plane's distance. A horizontal line would show the airplane's vertical position with respect to the correct glide path; *i.e.*, the horizontal line passes through the pip representing the plane when it is at the correct altitude.

The Teleran equipment is applicable to all but the smallest aircraft in its present form. Proponents of the system believe that television provides a means for relieving the pilot's present dependence on a maze of maps, meters, dials, scales, aural signals, etc. Pictorial presentation of information is almost effortlessly perceived and used.

59. Expected Developments. From the foregoing text, it will have become apparent that any system sufficiently comprehensive to provide complete navigational facilities and air traffic control must necessarily contain a basic minimum of functional and operating elements. The ideal system must be capable of providing the basic information for presentation to the aircrew and should also include means for automatic control of the aircraft. Full presentation of complete information on the azimuth, range, altitude, and identification of all aircraft within the control area must, of course, be available to the ground controllers. A difficult aspect of the over-all problem is the partially equipped airplane or one with defective apparatus. Intimately tied up with the problems of navigation and control is the hazard of collisions between aircraft and natural obstacles. The partially equipped airplane poses a real problem in this respect; the unequipped airplane does not belong in the air under adverse weather conditions.

It is evident that the present technical state of radio aids is such that the problems of either manual or automatic navigation, including take-off and landing, are virtually solved. On the other hand, the control of air traffic, including collision prevention, is far from solved. It would appear that the technical means for the solution of air traffic control are available but that much further development work is required. There are important psychological aspects of the control problem the resolution of which would prove of material aid to the system designers. Remaining as a heritage of the day when the pilot was virtually in complete control of his flight plan is the urge to present him with complete information for planning his flight. System designers, on the basis of analysis of the control problem, are fully cognizant of the fact that planning must be handled on the ground but are deterred from going to complete remote control from the ground by at least two important considerations. The first is the factor of the human life involved; the second is that radio has not yet fully demonstrated its capability of 100 per cent performance under all operating conditions. Thus air transports cannot be treated entirely as guided missiles, but some degree of freedom must be left to the initiative of the human pilot.

The net result of the contradictory requirements is that an integrated system must provide for presentation both at the ground station and in the aircraft, with just the proper balance between the two. Considerable practical experience with the most modern methods will be required before the exact degree of balance can be determined. The advantage for ground control is that elaborate computers may be employed to convert the basic (azimuth, distance, altitude, and identification) data to "future-position" data so that carefully considered flight instructions can be effected. It is apparent that rates of each of these factors, *i.e.*, velocity, acceleration, rate of azimuth bearing, rate of climb, etc., must be involved in such computations and that the task is too great for individual aircrews. The advantage for air presentation is that (at least from a collision viewpoint) the pilot is interested only in the aircraft about him and is perhaps in better strategic position to secure more accurate measurements.

The problem is not simple in either case and will in all probability be worked out on the basis of information gathered both in the aircraft and on the ground. For this reason, much emphasis cannot be ascribed to so-called "self-contained" air-borne systems, such as unassisted radar, except in so far as they integrate into a two-way over-all system.

Consider, for example, the instrumentation needed aboard the aircraft if it is to determine, without outside aid, all the data needed to prevent collision with other air-

craft and with natural obstacles.¹ With the present trend to higher speeds and since different airplanes may have different speeds, it is clear that a device showing only the range of other aircraft is entirely inadequate. If bearing data are added, thereby yielding present position of other aircraft, the information is still inadequate. The pilot needs information on future position of all aircraft of immediate interest to determine whether his present course is likely to bring him in collision and, if so, what maneuvers to take to avoid such collision. Future-position data require determination of the rate of change of present position for each neighbor airplane (referred to the pilot's own airplane) both as to speed and as to azimuth and elevation bearings. Even with the availability of such data, continuously presented to the aircrew, too detailed study may be required for practicable use. (It is to be noted that two airplanes flying a collision course, each at, say, 600 mph, will traverse 1 mile in 3 sec.) The net result is that automatic computers will be needed to give warnings and instructions as to evasive tactics. It should be clear, without actual appreciation of the details, that cooperation between the ground station and the aircraft crew (automatically effected) may yield an over-all system requiring much less equipment on the aircraft and, yet, simultaneously providing additional useful data at the ground station. For example, much of the data indicated to be needed for the self-contained air-borne instrument could be used by the ground controller for so scheduling traffic that the number of warnings to be issued by the air-borne system could be drastically reduced.

In addition to over-all systems planning, there are a number of urgently needed equipments which should reach the product stage in the near future. The need for surveillance of traffic on the airport runways and taxiing strips has already been indicated. A number of new devices should result from developments in this field.

In the field of radio altimetry, ICAO has stated that there are three major requirements yet to be met: (1) altimeters of moderate accuracy are required for pressure-pattern flying, (2) high-accuracy instruments are needed for reporting meteorological data, and (3) low-level altimeters are required for monsoon flying. It would obviously be desirable if a single instrument could be developed which would meet all three of these requirements.

One of the greatest deficiencies in modern aviation is a lack of adequate emergency and rescue facilities. ICAO has given recognition to this fact in recommending that the subject of practical rescue aids be studied by research and development organizations of the member states and has specifically recommended consideration of the following facilities for the use of long-range aircraft: (1) a transmitter operating in the m-f, h-f, and v-h-f bands, and a v-h-f band receiver, (2) automatic position-reporting equipment which derives its information from the long-range navigational aid and transmits on all three of the foregoing bands, (3) a DME beacon which can be located by aircraft provided with suitable equipment, and (4) a v-h-f pocket transceiver. These facilities should be capable of removal for use on a raft, or otherwise, and should be operable on an emergency battery and on hand generator equipment. ICAO believes that some locations will require the use of special search aircraft provided with a search radar set, DME facilities, and automatic d-f equipment.

Finally, there is an extremely attractive field of development in the application of electronic digital computers to the navigation and control problems of civil aviation. As the volume of air traffic increases, it is hardly conceivable that the present methods employed by ground controllers can cope with the control demands. The electronic digital computing machine has been applied to problems of amazing complexity and will find just such problems in civil aviation. It is conceivable that its aid will be required not only in actual traffic control but also in the flight-planning stages.

SPECIAL REQUIREMENTS IN AIR-BORNE INSTALLATIONS

Because of the special nature of the installation of radio navigational equipment aboard aircraft, special mechanical, electrical, and aerodynamic requirements are imposed. Reliability and simplicity of operation are essential. The equipment must

¹ See reference to Airborne Instruments Laboratory report at end of chapter.

be constructed to withstand continued vibration and landing shock without change in performance, and must operate under all conditions of weather encountered in flight. Space and weight must be kept to a minimum. The equipment must be capable of quick removal from the airplane for servicing or replacement. Simple but complete remote control of the equipment, including frequency change-over, etc., is essential. An adequate, efficient power supply is required. The antennas must be of sound electrodynamic and aerodynamic design. Special precautions are needed in eliminating various electrical disturbances arising on the airplane.

60. Radio Interference Sources. The electrical disturbances causing interference in aircraft radio receiving circuits arise from sources within as well as outside the aircraft. Interference sources within the aircraft are: (1) rotating machinery, such as generators, dynamotors, and motors, (2) electrical machinery with moving contacts, such as vibrators, speed governors, relays, and voltage regulators, (3) ignition systems for the main engine, auxiliary power unit, and ignition-type heaters, (4) pulsed electronic equipment and f-m devices, such as radar transmitters, pulse modulators, and radio altimeters, (5) internal background noise in receivers, due to tube hiss, thermal agitation, and shot effect.

Interference sources outside the aircraft consist of (1) atmospheric static, (2) precipitation static, which occurs when an aircraft passes through rain, snow, hail, ice crystals, or dust clouds, (3) man-made sources, such as concentrations of electrical equipment on the ground, as encountered over industrial areas, (4) ignition systems and other interference-producing equipment in nearby aircraft, and (5) undesired signals from other transmitters.

Prior to the Second World War, the lower gain of air-borne receivers then generally in use and the fact that those receivers seldom covered frequencies higher than 18 Mc, the comparatively small amount of rotating electrical machinery aboard the aircraft, and the absence of pulse-operated equipment, made the problem of suppressing locally generated interference less difficult than at the present time. Thorough shielding of the entire ignition system, from the magneto distributing heads up to and including the spark plugs, effectively eliminated ignition interference; the shielding of the various air-borne electrical devices, together with filtering their connecting leads, eliminated these as potential sources of interference; and the bonding of all metal cases and parts to a common ground formed by the aircraft fuselage eliminated the periodic discharge of static voltages accumulated on isolated metallic parts and components.

With the advent of the higher ignition voltages required for the operation of larger and faster aircraft at high altitudes and with the use of more sensitive receivers operating at higher frequencies, the interference-suppression problem again became acute. Exhaustive studies and tests by the Armed Forces and by manufacturers have pointed the way to a clearer understanding of the problems involved and have resulted in definite methods of treatment.

61. Elimination of Radio Interference. The suppression of ignition interference is now accomplished by shielding, by special receiver design, and by phasing-out of the interference. Proper shielding requires that the entire ignition system be enclosed in a continuous metallic shield, all parts of which have an extremely low r-f impedance not only between each other but also between themselves and ground. The selection of the proper shielding metal requires consideration of the effect of corrosion, as well as of the conductivity, weight, cost, and ease of fabrication. It has been found that, as the frequency is increased, the matter of fastening shielding cover plates becomes of greater importance, and that actual metal-to-metal contact is required in addition to the contact provided by the joining screws. Special receiver design consists of reducing the stray coupling which takes place to the receiver through internal antenna leads, power lines, remote-control cables, receiver cases, and headset cords or output leads. Sufficient reduction in such coupling can be achieved by careful design of equipment, the installation of filters, the shielding of leads, and the employment of interference-reducing techniques in the placing of open wiring to take full advantage of the shielding effect of the aircraft structure. At some of the lower frequencies, it is

possible to phase out the ignition interference by employing specially designed antennas and antenna-input circuits at the receiver.

In general, the interference from rotating machinery can be dealt with by means of filters of various types. In some cases simple by-pass capacitors connected to ground suffice; in other cases elaborate choke and capacitor combinations are required. In small machines, such filters may be connected in series with the power leads; in large machines, filters are connected between commutators, terminals, etc., and ground.

Sparking at relay contacts can be minimized by means of small capacitors or *RC* filters connected across the contacts. The most effective remedy in reducing the arcing in finger-type regulators is the use of *RC* filters.

The interference from pulsed equipment, such as radar, takes the form of an audio note which is the same as the pulse-repetition frequency. Even though ordinary communication receivers lie outside the band of the fundamental frequencies of radar transmitters, enough energy from the latter may enter the receiver to cause saturation of the first stage and subsequent rectification of the signal. Under these conditions, the pulse-repetition frequency, which is a modulation component of the radar signal, will be heard in the receiver. Rectification and consequent interference can also take place at faulty joints in the shielding case of the receiver or transmitter, at faulty joints in a transmission line, or at any place in the system where there happens to be a high r-f impedance across a joint. Many communication receivers have heterodyne oscillators which produce v-h-f harmonics. If the radar signal is mixed with one of these harmonics such that the resulting frequency is the same as that of the i-f stages of the receiver, the pulse-repetition frequency will appear in the receiver output.

Radiation of pulse energy can also occur from improperly shielded modulator cases and from pulse cables between the transmitters and modulators. Satisfactory reduction of radiation from the pulse cable calls for the use of double shielding, both shields being correctly terminated to ferrules attached to the fittings at either end of the cable. Pulse interference can also be conducted from the transmitter or modulator along the power wiring and control wiring. Low-pass filters are effective in conducting the undesired energy to ground.

Radar interference entering the receiver through the antenna or lead in may be reduced by means of network filters or wave traps tuned to the radar frequency. Other means of reducing radar interference have included blanking circuits, which cut off the receiver during the radar pulses, and the use of peak-noise limiters which reduce pulses of interfering energy of short duration but permit the reception of c.w. or voice.

The following installation practices have been evolved from experience for minimizing radio interference due to causes other than precipitation static. These practices cover (1) location of air-borne equipment, (2) bonding of equipment, (3) location of antenna and leadin, (4) routing of electrical, radio, and radar cables, and (5) installation and bonding of metallic conduit.

Location of Equipment. Communication receivers and transmitters should be so located that their antenna lead-in wiring is as short as possible, preferably not over 1 ft in the case of the receiver. Dynamotors, generators, electric motors, etc., should be located under decks, behind bulkheads, and in positions remote from receivers, to prevent interference fields radiated from their cases from coupling to the receiver lead in wiring. All rotating electrical machinery should be located away from the openings in the aircraft's structure to prevent interference fields from such equipment from radiating directly to external antennas. Radar modulators and transmitters should be located in different compartments from those housing communication receivers. Auxiliary power units should be located away from openings in the fuselage or hull, so that ignition interference from the unit will not radiate to external antennas through the openings.

Bonding of Equipment. The cases of receivers, transmitters, and radar units, and the main frames of dynamotors, generators, etc., should be bonded to the aircraft structure with conductors having a resistance not exceeding 0.0025 ohm. The cases of induction vibrators or booster coils should be bonded to engine mounts with conductors of the same order of resistance. All bonding jumpers on electrical and electronic equipment should not exceed 4 in. in length.

Location of Antennas and Leadins. Main communication antennas should be so installed that the leadin insulator will not be more than 1 ft away from the receiver antenna post. Main communication antennas or external leadins should be installed away from engine cowl openings; and antenna leadins should be located at a distance from electrically controlled propellers. Main communication antennas should be located outside the field of scanning radar antennas. Loop antennas should be located away from engines, to avoid ignition interference pickup; and should also be so located that they are not affected by direct radiation from radar antennas. V-h-f antennas should be so located as to minimize the effects of propeller modulation and ignition interference, and should be shielded from radar antennas by the aircraft structure. Antenna leadins inside the aircraft should be as short as possible. They should be routed near the aircraft structure to avoid the formation of loops of large area with the structure; and the shielding effect of structural members should be utilized to full advantage in preventing coupling to electrical wiring loops.

Routing of Cables. Power cables should be routed at a distance from the leadin wiring and from openings in the airplane structure so that they cannot radiate interference to the external antenna. Radar pulse cables and transmission lines should have their shielding correctly terminated to ground at plug connections; and radar pulse cables should be routed separately from all other wiring and at least 1 ft away therefrom. Power and control cables carrying energy at high-interference levels should be routed separately from leads carrying energy at low-interference levels. The shielding effect of the aircraft structure should be utilized to full advantage in routing power and control cables, to prevent the inductive coupling of interference into other connecting leads. Ignition-switch leads should be routed in metallic conduit throughout their entire length from magnetos to ignition switches; and ignition leads from magnetos to booster coils or induction vibrators should be similarly shielded.

Installation and Bonding of Metallic Conduit. In all-metal aircraft, it is feasible to obtain satisfactory radio performance without resorting to the use of metallic conduit shielding, except as follows: (1) ahead of the fire wall and throughout the ignition system, (2) where the natural shielding of the aircraft's structure is inadequate or impracticable to use, and (3) where such conduit is needed for mechanical support, protection, or ease of wiring replacement. The usual method of bonding flexible conduit by means of a clamp and bonding jumper has been found unsatisfactory for interference reduction. This is because the clamp compresses the conduit with a high unit pressure around a limited area of contact and, after vibration in service, the clamp loosens and produces high resistance or intermittent contact with the conduit. A preferred method is to use a flared split sleeve under the standard clamp, thus exerting a low unit pressure over a large area. The split sleeve should be soldered to the conduit whenever the material permits.

62. Reduction of Precipitation Static. Methods of various degrees of effectiveness have been devised for coping with the problem of precipitation static. One of the earliest methods, consisting of the use of a shielded loop antenna, reduced such static to a limited degree. Other methods, which were somewhat more effective, employed systems of trailing wires to provide a discharge point well removed from the aircraft antennas, and controlled the discharge by means of a resistor so that oscillating corona was minimized. One such system employed a very fine wire only a few inches long, which was trailed from the tail of the aircraft and was kept at the desired tautness by means of a tiny parachute attached to the end. The inner end of the wire was connected to the aircraft structure through a suppressor resistance. With this arrangement, the potential gradient at the wire created ionization and caused partial dissipation of the aircraft's charge. This method not only produced an inadequate discharge but also proved mechanically impracticable. Other similar methods employed trailing wires at the ends of cords, the wires extending into the slip stream from the tail of the airplane, and the cords having high electrical resistance for the purpose of suppressing impulse oscillations inherent in such types of ionizers.

Wet- and Dry-wick Static Dissipators. Older systems, which proved of little practical value, were succeeded first by wet-wick systems, then by dry-wick dissipators. The wet wick, which proved unsatisfactory for various electrical, mechanical, and maintenance reasons, consisted of cotton rope-floss wicks attached to the trailing edges of the wings and to the tail. Glycerin treated with salt was allowed to drain from

small reservoirs into the wicks, thus lowering the electrical resistance of the cotton fibers, which would otherwise act as insulators, and allowing ionization at their ends.

Dry-wick precipitators, developed as the result of an intensive investigation by joint action of the Army and Navy, consist of a rope-floss wick of fine cotton fibers rendered partially conductive by treatment, during manufacture, with a colloidal solution of silver salt. The rope floss is encased in a flexible plastic tube having an outside diameter of about 0.25 in. and projects about 1 in. from the outer end of this tube. The other end of the wick and tube is secured in a conventional electrical lug which is attached to the trailing edge of the wing. As in the case of former systems, the resistance of this wick is of the order of several megohms to introduce an oscillation suppressor between the free end (where ionization takes place) and the metal structure of the aircraft. The fibers of this wick are torn loose by the action of the slip stream and thus require frequent trimming and eventual replacement of the plastic tube. The sponsors of the dry wick have recommended the installation of two or three wicks at the trailing edges of each wing (near the tips) of each stabilizer and of the rudder. Despite the claims made for this device, other investigators have concluded that the system is ineffective in controlling radio interference. Furthermore, one investigator reports that qualitative tests on a considerable number of dry wicks have disclosed the fact that these wicks experience serious resistance aging. Such aging increases the resistance to many times the initial value and correspondingly decreases the effectiveness of the wicks as static dischargers.

Metal-bristle Ionizers. A still more recent system, represented as superior to any previous method, makes use of an ionizer consisting of several tufts of 2-mil spring-brass bristles having points chemically etched to vanishing fineness, anchored through holes in a narrow rectangular brass plate. This plate is bent over the edge of the airfoil and is riveted to the latter so that the bristles project outward into the slip stream about 1.5 in. Heavier, shorter, and stiffer bristles with similarly sharpened ends are used in units mounted transversely to the air flow at the tips of the wings and stabilizers, and at the top of the rudder. By installing an adequate number of bristle ionizers at electrically strategic locations, high-potential gradients are prevented from developing on parts of the airplane. It is claimed that tests on laboratory aircraft under conditions of snow and other forms of precipitation have demonstrated that these ionizers neither develop radio noise nor permit it to develop from parasitic sources. Cadmium plating prevents corrosion and weathering of these ionizers and hence ensures long life with a negligible amount of maintenance.

Controlled Dischargers Utilizing High D-c Voltage. A proposed method of dealing with the problem of snow static, which is generally recognized as the most serious form of precipitation static on aircraft, gives promise of being highly effective, especially in the case of high-speed aircraft. This new method, which evolved from flight research conducted both during and before the war, is based upon the theoretical desirability of discharging from an aircraft all electric charges the instant they are collected, thus maintaining the total charge on the aircraft at zero. The proposed method consists essentially of producing a noninterfering¹ negative-corona discharge from sharp points in the slip stream adjacent to the wings and the tail surfaces. This is accomplished by applying a high d-c voltage, of the order of 5 kv, to special discharge points mounted either at the trailing edges or above and below the surfaces. With this arrangement, the d-c voltage maintains a continuous corona discharge from the discharge point to the aircraft structure. The flow of air created by the movement of the aircraft blows away the ions from the corona discharge, thus causing a flow of current from the aircraft and a diminution in the total charge on the aircraft structure. This discharge current increases with the velocity of the air flow and hence with an increase in the speed of the aircraft. On the larger aircraft, several discharge points could be supplied from the same rectifier, and a separate bank of dischargers could be located at each of the main extremities, such as the wing tips and the edges of the horizontal and vertical tail surfaces. On very high-speed aircraft, this system of controlled dischargers offers

¹ Proponents of the system state that this is feasible according to basic theoretical considerations.

the possibility of allowing the radio operator to switch into operation additional units as static-forming conditions increase in severity.

One proposed form of discharge point consists of a conical supporting insulator about 0.5 in. in diameter, at the end of which, and perpendicular to its axis, is mounted a small stainless-steel rod. The latter is bent backward toward the supporting insulator in the form of an arc, and its ends are provided with sharp points which terminate about 0.5 in. from the surface of the aircraft structure. The high-voltage d.c. is supplied to the steel bar by means of a connecting lead which passes axially through the insulator; and this d-c voltage can be obtained from a small rectifier operating directly from a 60-cycle, a-c, low-voltage source.

References

Books:

- Bond, D. S.: "Radio Direction Finders," McGraw-Hill, 1944.
 Fassbender, H.: "High-frequency Technique in Aircraft," Springer, 1932.
 Morgan, H. K.: "Aircraft Radio," Pitman, 1939.
 Redpath, P. H., and J. M. Coburn: "Air Transport Navigation," Pitman, 1943.
 Roberts, H. W.: "Aviation Radio," Morrow, 1945.
 Sandretto, P. C.: "Principles of Aeronautical Radio Engineering," McGraw-Hill, 1942
 Stewart and Pierce: "Marine and Air Navigation," Ginn, 1944.
 United States War Department: "Report of Electronic Subdivision Advisory Group on Air Navigation," Air Materiel Command, Engineering Division, Army Air Forces, Wright Field, Dayton, O. (Rpt. TSELCS-SP2), June, 1946.

Publications and Technical Papers:

- Airborne Instruments Laboratory, Inc.: "An Analysis of the Problems of Aircraft Collision Prevention," Rpt. 508-1, September, 1946.
 Beach, R.: Electrostatic Ills and Cures of Aircraft, *Elec. Eng.*, April and May, 1947.
 Bennett, W. H.: "Elimination of Precipitation Static," paper presented at IRE 1947 National Convention, New York Mar. 4, 1947; unpublished as of June 1, 1947.
 Busignies, H. and P. R. Adams: "First Phase of Development and Flight Testing of Navar System of Air Navigation and Traffic Control Aids," revised title of paper presented at IRE 1947 National Convention, New York, Mar. 5, 1947; unpublished as of June 1, 1947.
 ———, and M. Dishal: "Relations between Bandwidth, Speed of Indication and Signal-to-noise Ratio in Radio Navigation and Direction Finding," paper presented at IRE 1947 National Convention, New York, Mar. 4, 1947; unpublished as of June 1, 1947.
 Caporale, P.: The C.A.A. Instrument Landing System, *Electronics*, February, 1945, pp. 116-124; March, 1945, pp. 128-135.
 Carson, V. S., S. L. Seaton, M. Rothman, and M. Pomerantz: "Low-frequency Loran, Watson Laboratories, Air Materiel Command, Army Air Forces," paper presented at IRE 1947 National Convention, New York, Mar. 4, 1947; unpublished as of June 1, 1947.
 Council for Scientific and Industrial Research: "The Australian Multiple Track Radar Range," undated, prepared for purposes of demonstration to PICAO at Ottawa, Ontario, Canada, September to October, 1946.
 ———: "A Multiple Track Radar Range," *Conf. Paper RPR 6*, May, 1946.
 Crichlow, W. Q.: "The Comparative Accuracy of Various Existing and Proposed Radio Navigation Systems," *Nat. Bur. Standards, Rpt. CRPL-4-1*, Dec. 30, 1946.
 Darwin, F. A.: "A Comparison of Interrogation by Search Radars and by Separate Interrogators in Pulse Transponder Systems," paper presented at IRE 1947 National Convention, New York Mar. 4, 1947; unpublished as of June 1, 1947.
 Federal Tele-Communication Laboratories, Inc.: "Navar—Flexible Coordinated Series of Radio Aids to Air Navigation and Traffic Control," descriptive write-up, 1947.
 Hazeltine Electronics Corp.: "Hazeltine Distance-measuring Equipment for All-weather Navigation," *Tech. Rpt. 2044-W*, Sept. 25, 1946.
 ———: "Hazeltine Lanac System for All-weather Flying," *Tech. Rpt. 2043-W*, Sept. 25, 1946.
 ———: "Hazeltine PICA0 Demonstrations at Indianapolis, Oct. 10-23, 1946," *Tech. Rpt. 2045-W* Oct. 30, 1946.
 Jensen, R. C., and R. A. Arnett: Airborne Radar for Navigation and Obstacle Detection, *Elec. Eng.*, May, 1946, pp. 307-313.
 Lyman, J.: "The Application of Microwaves to Instrument Landing of Airplanes," paper presented at the National Air Transport Meeting, Washington, Oct. 24, 1946; published by Institute of the Aeronautical Sciences, New York.
 ———: "Coordinated Air Traffic Control," paper presented at the Indianapolis meeting of the Provisional International Civil Aviation Organization, Oct. 7-26, 1946.
 ——— and G. Litchford: "The Application of Microwaves to the Guidance and Control of Aircraft," paper presented at IRE 1947 National Convention, New York, Mar. 5, 1947; unpublished as of June 1, 1947.

- McIlwain, K.: "Hazeltine Lanac System—Laminar Air Navigation and Anti-collision," paper presented at IRE 1947 National Convention, New York, Mar. 5, 1947; unpublished as of June 1, 1947.
- Pierce, J. A.: "The Basic Characteristics of Radio Aids to Navigation," paper presented at the symposium on the Guidance and Control Aspects of Guided Missile Development, at M.I.T., February, 1947; submitted to IRE for publication; unpublished as of June 1, 1947.
- Provisional International Civil Aviation Organization: "Final Report of the First Session, Special Radio Technical Division (COT)," *Doc. 2553-COT/26*, Montreal, Canada; report dated January, 1947.
- : "Report of the Second Session, Aeronautical Telecommunication and Radio Aids to Air Navigation Division (COM)," *Doc. 2495-COM/163*, Montreal, Canada; report dated December, 1946.
- Radio Corporation of America: "Teloran for Air Navigation and Traffic Control," descriptive write-up, Sept. 1, 1946.
- Robertson, S. D.: "Targets for Microwave Radar Navigation," paper presented at IRE 1947 National Convention, New York, Mar. 4, 1947; unpublished June 1, 1947.
- Sperry Gyroscope Company, Inc.: "Coordinated Air Traffic Control," descriptive write-up, *Eng. Div. Pub. ED 101*, January, 1946.
- : "Sperry Loran," descriptive bulletin, May, 1946.
- U.S. Coast Guard: "Electronic Navigational Aids (Loran—Radar—Racon)," Public Information Division, Washington, 1945.
- U.S. Department of Commerce: "Advancing Air Navigation—The CAA Program," CAA Aviation Information, CAA, 1946.
- : "Airborne Radar for Collision Warning, Technical Development Service, CAA, October, 1946," prepared for purposes of demonstrations to PICAO at Indianapolis, Oct. 7–26, 1946, undated.
- : "The CAA Radio Instrument Landing System with Simultaneous Voice, Technical Development Service, CAA, October, 1946," prepared for purposes of the demonstrations to PICAO at Indianapolis, Oct. 7–26, 1946, undated.
- : "The CAA-RTCA Instrument Landing System," Part I, Development and Installation, *Technical Development Rpt. 35*, CAA, October, 1943; Part II, Tests and Modifications, *Preliminary Technical Development Rpt. 36*, CAA, October, 1943.
- : "The CAA VHF Omni-directional Radio Range with Simultaneous Voice, Technical Development Service, CAA, October, 1946," prepared for purposes of the demonstrations to PICAO at Indianapolis, Oct. 7–26, 1946, undated.
- : "Classification of CAA Communications Aids," 1st ed., Federal Airways Manual of Operations 1-FA-3, CAA, Oct. 1, 1946.
- : "Digest of Airways Operations Service," Office of Federal Airways, CAA, August, 1946.
- : "A Digest of the Programs and Appropriations for Air Navigation Facilities, 1927–1947," Air Navigation Facilities Service, CAA, undated.
- : "Flying the Instrument Landing System," Airways Flight Inspection Staff, CAA First Region, New York, CAA, January, 1947.
- : "Radar Aid for Airport Traffic Control; Technical Development Service, CAA, October, 1946," prepared for purposes of the demonstrations to PICAO at Indianapolis, Oct. 7–26, 1946, undated.
- : "Radio Pulse Distance Measuring Equipment, Technical Development Service, CAA, October, 1946," prepared for purposes of demonstrations to PICAO at Indianapolis, Oct. 7–26, 1946, undated.
- : "U.S. Air Navigation Aids—A Brief Description, CAA," prepared for purposes of the demonstrations to PICAO at Indianapolis, Oct. 7–26, 1946, undated.
- U.S. Navy Department: "Elimination of Radio Interference in Aircraft," Bureau of Aeronautics, *Rpt. Navaer 16-5Q-517*, Mar. 1, 1946; restricted publication.
- : "Loran Transmitting Station Manual," Electronics Division, Bureau of Ships, *Rpt. Navships 900.060 A*, Mar. 1, 1945; officially declassified by Navy Department, February, 1946.
- U.S. Senate: "Investigation of Air Safety, Interim Report of the Committee on Interstate and Foreign Commerce," *Rpt. 36*, Feb. 22, 1947.
- U.S. War Department: "War Department Technical Manual TM 11-243," Radio Set SCR-291-A (direction finding), Mar. 24, 1944; declassified publication.

Suggested Reading:

- Adams, P. R., and R. I. Colin: Frequency, Power and Modulation for Long-range Navigation Systems, *Elec. Commun.*, **23**, 144–158, June, 1946.
- Aircraft Radio Corporation: "Instruction Books for Type 15 VHF Omni-directional Range Receiving Equipment," Feb. 21, 1947.
- Alford, A., and A. G. Kandoian: "Ultra-high-frequency Loop Antennas," *AIEE Tech. Paper* 40–45, January, 1940.
- Busignies, H., P. R. Adams, and R. I. Colin: Aerial Navigation and Traffic Control with Navaglobe, Navar, Navaglide and Navascreen, *Elec. Commun.*, **23**, 113–143; June, 1946.
- Buttner, H. H., and A. G. Kandoian: Development of Aircraft Instrument Landing Systems, *Elec. Commun.*, **22** (3), 179–192, 1945.
- Colin, R. I.: Survey of Radio Navigational Aids, *Elec. Commun.* **24** (2), 219–261, June, 1947.
- Council for Scientific and Industrial Research: "The Australian Distance Indicator," *Conf. Paper RPR7*, May, 1946.

- Diamond, H.: The Cause and Elimination of Night Effects in Radio Range Beacon Reception, *Nat. Bur. Standards J. Research, Research Paper* 513, 10, 7-34, 1933; *Proc. IRE*, 21, 808-832, 1933.
- : Performance Tests of Radio System of Landing Aids, *Nat. Bur. Standards, J. Research, Research Paper* 602, 11, 463-490, October, 1933.
- and F. W. Dunmore: A Radio Beacon and Receiving System for Blind Landing of Aircraft, *Nat. Bur. Standards, J. Research, Research Paper* 238, 6, 897, 1930; *Proc. IRE*, 19, 585, April, 1931.
- and F. G. Gardner: Engine Ignition Shielding for Radio Reception on Aircraft, *Nat. Bur. Standard, J. Research, Research Paper* 158, 4, 415-424, March, 1930; *Proc. IRE*, 18, 840-861, May, 1930.
- Dieckman, M.: "Report on the Status of a Direct-indicating Airplane Radio Direction Finder," Deutsche Versuchsanstalt für Luftfahrt, Berlin, July, 1927.
- Engel, F. H., and F. W. Dunmore: A Directive Type of Radiobeacon and Its Application to Navigation, *Nat. Bur. Standards, Sci. Paper* 840, 1923.
- Engel, J. S.: Landing Aircraft with Ground Radar, *Elec. Commun.*, 25, 72-81, March, 1947.
- Espenschied, L., and R. C. Newhouse: A Terrain Clearance Indicator, *Bell System Tech. J.*, 18, 222, January, 1939.
- Haller, G. L.: Constants of Fixed Antennas on Aircraft, *Proc. IRE*, 26, 415, April, 1938.
- Hinman, W. S., Jr.: A Radio Direction Finder for Use on Aircraft, *Nat. Bur. Standards, J. Research, Research Paper* 621, 11, 733-741, 1933.
- Holloman, G. V.: Automatically Controlled Blind Landing, *SAE J.*, 42, 13, June, 1938.
- Jackson, W. E., and D. M. Stuart: Simultaneous Radio-range and Telephone Transmission, *Proc. IRE*, 25, 314, March, 1937.
- Kear, F. G.: Phase Synchronization in Directive Antenna Arrays with Particular Application to the Radio Range Beacon, *Nat. Bur. Standards, J. Research, Research Paper* 581, 11, 123-139, 1933.
- McKeel, P. D., J. M. Lee, and H. I. Metz: "Development of an Improved U-h-f Radio Fan Marker, Rpt. 13, Safety and Planning Division, CAA, July, 1938.
- Murphy, W. H.: Space Characteristics of Antennae, *J. Franklin Inst.*, 201, 411, 1926; 203, 289, 1927.
- PICAO: "Rules of the Air and Traffic Control," *Doc. 2010-RAC/104*, February, 1946.
- Sandretto, P. C.: Some Principles in Aeronautical Ground-radio Station Design, *Proc. IRE*, 27, 5-11, January, 1939.
- Third Commonwealth and Empire Conference on Radio for Civil Aviation, Summer, 1945.
- U.S. Coast Guard: "Electronic Navigational Aids," *Immran Doc. 43*, New York-New London, April, 1947.
- U.S.: Department of Commerce: "Army-Navy-CAA Standards for the Control of Instrument Flight Rule Traffic"; Apr. 1, 1946. "A Digest of the Programs and Appropriations for the Establishment of Air Navigation Facilities," CAA, July 7, 1947.
- U.S. War Department: "The Range, Reliability and Accuracy of a Low-frequency Loran System," *Rpt. ORS-P-23*, January, 1946.
- White, W. D.: "The Function of Air Traffic Control," paper presented at IRE 1947 National Convention, New York, Mar. 5, 1947; unpublished as of June 1, 1947.

CHAPTER 22

CODE TRANSMISSION AND RECEPTION

BY JOHN B. MOORE,¹ AND DAVID S. RAU¹

1. **Radio communication**, as distinguished from radio broadcasting of educational and entertainment programs, is carried on chiefly by means of some one of the recognized telegraph codes. Radiotelegraph signals are, therefore, made up of short and long periods of constant signal strength separated by idle periods of proper duration to correspond to the combinations of dots, dashes, and spaces comprising the characters of the code being used. The design of the entire system must be such that the lengths of the dots, dashes, and spaces in the signal supplied to the receiving operator are substantially the same as they were made by the transmitting operator. In a simple system operated at slow speeds no special difficulties are encountered in meeting this requirement. Present-day commercial systems, however, which utilize remote control from a central traffic office and which are operated at high keying speeds, impose severe requirements on all the equipment used.

STANDARD CODES

2. **The International Morse code** (also called Continental code) is widely used for international communication. This consists of dots and dashes as depicted in Fig. 1. Specially marked and accented letters such as are used in German, French, and the Scandinavian languages have special characters which are used when working a station in the same country or its possessions. When communicating with a foreign station, these letters are either replaced by a combination of unaccented letters or in some cases the unaccented letter is transmitted alone. Some countries such as Japan and Egypt, having alphabets differing radically from the Latin alphabet, use special codes for working within the country or to their own ships. Nationals of such countries desiring to transmit a message in their own language to a foreign country must spell out the sounds of their own words in one of the languages using the Latin alphabet.

3. **Printing-telegraph Codes.** Various types of printing systems in which the received signal is automatically printed in standard letters on a paper tape or roll are being employed on the high-grade radio circuits of the world. The trend is toward such reception as a substitute for manual transcription. Codes employed are the International Morse code, the "five-unit" Baudot code, and the recently developed "seven-unit" code.²

The Baudot code consists of a different combination of marking and spacing impulses for each character, there being a total of 32 possible combinations for the five periods utilized. One of these combinations is used for operating the printer mechanism to upper case, and another to restore it to lower case. Consequently 60 other characters or operations may be transmitted. Generally most administrations utilize the same combinations for lower-case characters so that the printer of one may operate into that of another, but upper-case character arrangements differ to suit the particular requirements of the user. Several variations are indicated in Fig. 2.

The "seven-unit" printer code divides the time allotted for transmission of any one code group into seven consecutive and equal time intervals. Only three of these possible seven pulses are used for any code group or character. The receiving equipment is so designed that a received group containing fewer or more than three pulses or

¹ Engineering Department, RCA Communications, Inc.

² U.S. Patent 2183147.

A ---	Period..... : ---
B ----	Comma..... , ---
C ----	Colon..... : ---
D ---	Question mark, or request for repetition of a transmission not understood..... ? ----
E ·	Apostrophe..... ' ---
F ----	Dash or hyphen..... - ----
G ----	Fraction bar..... / ----
H ----	Parenthesis (before and after words) () ----
I ··	Underscore (before and after words or part of sentence)..... _ ----
J ----	Equal sign..... = ----
K ---	Understood..... ···
L ----	Error..... ·····
M ---	Cross or end-of-telegram or end- of-transmission signal..... ···
N ---	Invitation to transmit..... ---
O ----	Wait..... ·····
P ----	End of work..... ·····
Q ----	Starting signal (beginning every transmission)..... ---
R ----	Separation signal for transmission of fractional numbers (between the ordinary fraction and the whole number to be trans- mitted) and for groups con- sisting of figures and letters (between the figure-groups and the letter-groups)..... ···
S ···	
T ---	
U ----	
V ----	
W ----	
X ----	
Y ----	
Z ----	
· (German) ----	
Á or À (Spanish-Scandi- navian) ----	
CH German-Spanish	
É (French) ----	
Ñ (Spanish) ----	
Ö (German) ----	
Ü (German) ----	
1 ···	
2 ···	
3 ···	
4 ···	
5 ···	
6 ···	
7 ···	
8 ···	
9 ···	
0 ----	

FIG. 1. The Continental code.

marking intervals will cause a special "error sign" to be printed instead of an incorrect letter or figure.

4. **Business Codes.** Business concerns that have a large volume of telegraph communication use so-called "five-letter" or "ten-letter" codes. Standard codes for such use are available and consist of groups of letters arranged alphabetically; each group

UPPER CASE					LOWER CASE	5-unit				7-unit *						
Weather (U.S.)	AT & TCo. TWX	U.S. Army †	R.C.A.	Inter-nat'l.		1	2	3	4	5	1	2	3	4	5	6
↑	-	-	-	-	A	o	o				o		o	o		
⊕	5/e	?	?	?	B	o		o	o		o	o				o
o	1/8	:	:	:	C		o	o	o		o			o	o	
/	§	§	§		D	o		o			o			o	o	
3	3	3	3	3	E	o								o	o	o
→	1/4	!	!		F	o		o			o			o	o	
↘	⊗	⊗	⊗		G		o	o	o						o	o
↓	stop	⊥	*		H			o	o			o			o	o
8	8	8	8	8	I		o	o			o			o	o	
↙	,	,	bell	bell	J	o	o	o				o	o	o	o	
←	1/2	(((K	o	o	o					o	o	o	o
↖	3/4)))	L		o		o			o	o		o	
.	M			o	o		o	o	o	o	o	
⊙	7/8	,	,	,	N			o	o			o	o	o	o	
9	9	9	9	9	O			o	o	o	o	o	o	o	o	
o	o	o	o	o	P		o	o	o			o	o	o	o	
!	!	!	!	!	Q	o	o	o	o			o			o	o
4	4	4	4	4	R		o	o					o	o	o	
bell	bell	bell	!	!	S	o	o	o				o	o	o	o	
5	5	5	5	5	T				o	o	o	o	o	o	o	
7	7	7	7	7	U	o	o	o			o	o			o	
⊖	3/8	;	;	*	V		o	o	o	o			o	o	o	o
2	2	2	2	2	W	o	o		o	o	o	o	o	o	o	o
/	/	/	/	/	X	o	o	o	o				o	o	o	o
6	6	6	6	6	Y	o	o	o	o	o	o	o	o	o	o	o
+	"	"	"	+	Z	o	o		o	o	o	o	o	o	o	o
space	space	space	space	space	space		o				o	o	o			
ltrs	ltrs	ltrs	ltrs	ltrs	ltrs	o	o	o	o	o		o	o	o		
figs	figs	figs	figs	figs	figs	o	o	o	o	o	o	o	o	o	o	
c.r.	c.r.	c.r.	c.r.	c.r.	c.r.			o			o	o	o	o	o	o
l.f.	l.f.	l.f.	l.f.	l.f.	l.f.		o				o	o	o	o	o	o

c.f. = carriage return
 l.f. = line feed

* 7-unit printer uses RCA standard keyboard.
 † Also used by AT&TCo (Bell System) private line teletype nets

FIG. 2. Printer key arrangements and codes.

standing for a complete sentence or part of a sentence. Special and private codes are also used. Large concerns often have a department for the coding and decoding of coded telegraphic messages.

SIGNAL CHARACTERISTICS

5. **Character Formation.** The smallest time unit used in code characters is the dot. Based upon automatic Morse transmission the time required to transmit one dot

includes the duration of the space separating the dot from the next element of the character. As the duration of the dot itself and the following space are equal, they constitute a cycle. It is customary then to speak of keying speeds in "dots per second," "dot-cycles per second," or just "cycles per second." Alternatively, keying speed may be expressed in "bauds," the baud being equal to one-half a dot-cycle per second.¹

In the Baudot code used for printing telegraph the duration of the character is divided into five equal periods, for any one of which either a marking or a spacing (no current or reverse current) impulse may be transmitted. In the more commonly used nonsynchronous equipment one additional impulse is required ahead of the character to start the mechanism in operation and another at the end of the character to allow the mechanism to clear for the next character and also provide time for the receiving machine to print the identified character. The total number of elements per character in this system is therefore seven. The space between words is equivalent to a full character including the start and stop impulses. For calculation of keying frequency, the unit element corresponds to the marking portion of a dot in the Morse code. This is one-half dot-cycle, and the corresponding unit of speed is one-half dot-cycle per second or one baud as with the Morse code. With the seven-unit start-stop equipment operating at a nominal speed of 60 wpm, each of the character impulses has a duration of approximately 22 millisecc. The reciprocal of this value gives the keying speed as 23 dot-cycles per second, or 46 bauds.

6. Required Frequency Range. A square-wave shape such as a succession of dots, where the value of the current or voltage rises instantly to a steady value at which it remains for one half cycle and then instantly drops to zero, can be analyzed into the fundamental and all its odd harmonics. The equation of the voltage wave is

$$e = \frac{4E}{\pi} \left(\sin x + \frac{1}{3} \sin 3x + \frac{1}{5} \sin 5x + \dots \right) \quad (1)$$

which holds for values of x between $-\pi$ and $+\pi$. For most practical telegraphic purposes, it is only necessary for the system to pass the fundamental, third, and fifth in their proper intensity and phase, as terms of higher order do not add sufficiently to the fidelity to warrant building the equipment to handle them. The frequency range required by a sufficient number of higher order harmonics to give appreciable improvement can often be used to better advantage for additional channels.

For any service where the received signal strength rises to the same maximum value on every dot and dash, it is not necessary to pass even the third harmonic of the keying frequency. A system that will pass the second harmonic of the fundamental keying frequency is satisfactory. The receiving equipment can be adjusted to operate at a fairly definite level on the building up and decaying of the current or voltage wave so as to give characters which are neither too heavy (long) nor too light (short) as compared to the spaces. However, in a system where the received signal may vary by 2:1 or more in intensity at fairly short and frequent intervals, it is necessary to have quite a steep rise and fall of the received signal at make and break to obtain a constant "weight" of keying. This applies particularly to automatic reception, where the signal operates a recording device either directly from amplifiers or through a relay of either the mechanical or vacuum-tube types. For aural reception it is desirable to retain the harmonics of the keying frequency, as the signal then sounds cleaner cut and more definite, making it easier to read.

Cases of interference, in both the radio and the landline portions of a system, are

¹ By Opinion 301 of the International Consulting Committee on Telegraphy (CCIT) at its meeting in Warsaw in October, 1936, the "baud" was officially defined as "the unit of telegraphic speed or of rapidity of modulation corresponding to one minimum element per second." For lack of a term to represent the minimum element itself, some writers have used the expression "baud element" or just plain "baud" in lieu of the unwieldy expression "half dot-cycle." References have consequently been made to "bauds per second" which expression, though officially incorrect, has come to be accepted through common usage as proper. It is desirable that agreement be reached on use of the term "baud" as it has been officially defined, but it may be necessary to help bring this about by adopting a new term to represent the elemental unit of code.

sometimes encountered where it is necessary slightly to round off the sharp, square envelopes of the dots, in order to reduce or eliminate the interference or cross talk caused by the too sudden rise and fall of current.

Where the exact effect of a given circuit on the shape of a square input wave is desired, the range of frequencies passed by the system must be considered as a continuous band rather than dealing with only odd harmonics of the keying frequency.

The usual modulation and side-band theory of radiotelephony is applied to code transmission by considering the fundamental keying frequency and such of its harmonics as are passed, to modulate the carrier 100 per cent. The total band width required to be passed by the entire system is equal to twice the frequency of the highest harmonic of the keying speed that it is desired to retain.

7. Speeds Attainable. Speeds of transmission range from about 15 up to 500 wpm; the corresponding keying frequencies being 6 to 200 dot cycles per second. Work with ships and with aircraft is carried on mainly at speeds up to about 35 wpm; transmission is by means of a manually operated telegraph key, reception is by ear. In point-to-point service, such as transoceanic, traffic speeds normally range from 30 up to 300 wpm depending upon the type of equipment used, transmission conditions, and the amount of traffic to be handled. Keying is done by machine almost entirely, hand-operated keys being used only for minor service communications. Reception is generally by means of a printer or an ink recorder, the telegraphic characters on the tape of the latter being transcribed on a typewriter by the operator. Aural reception is resorted to only under adverse conditions. In radio systems where multiplex equipment is employed on the circuits, each channel of the four usually going over a single circuit will operate at approximately 60 wpm. This gives the circuit a total capacity of 240 wpm. Eight 60-wpm channels may be operated on circuits of high quality.

8. Fidelity of the mark-to-space ratio, although important at all speeds,¹ requires special attention when automatic operation at speeds in excess of 100 wpm is to be maintained. Where the duration of the mark portion of a dot is only 12 millisecond or less, factors that are disregarded at slow speeds become of primary importance. Automatic transmitters, relays, and electrical circuits should be fast enough so that the signal supplied to the recording equipment will not be heavier than 60:40 or lighter than 40:60 in mark-to-space ratio at the highest speed used. At 200 wpm, which is not exceptional in present-day h-f work, this means a variation of not more than 1.25 msec in the duration of a dot. While it is sometimes possible to compensate for heavy or light keying characteristics by means of relay adjustments in another portion of the system, this should not be depended upon for obtaining the desired over-all fidelity. Each unit of the system should be capable of giving the required fidelity at a speed in excess of the maximum operating speed, the margin required depending on the number of elements in the over-all system and the fidelity of each.

9. Multiplex operation over a radio circuit has certain very definite advantages for traffic handling. No single operator can keep traffic moving at 100 wpm. The radio circuit, however, is often capable of handling several times this speed. Economical operation then requires that two or more operators be assigned to the circuit. Multiplex equipment permits doing this in the most expeditious and straightforward manner since each operator then has a channel under his complete control. This makes it possible to use suitable printing-telegraph equipment efficiently, each such channel being handled by a single operator at a speed of approximately 60 wpm. Four such channels give a circuit capacity of 240 wpm, with no complications such as are experienced when such high-speed operation is attempted over a single-channel circuit employing tape transmission and reception.

Two basic types of multiplex have been employed. One utilizes two or more modulating frequencies which are applied to the radio transmitter. The other employs the time-division principle.² The chief advantage of the latter is that it can be applied

¹ Rea, W. T., Effect of Telegraph Distortion on the Margins of Operation of Start-stop Receivers, *Bell System Tech. J.*, July, 1944.

² Callahan, J. L., Mathes, R. C., and Kahn, A., Time Division Multiplex in Radiotelegraph Practice, *Proc. IRE*, January, 1938.

to any radiotelegraph circuit which will properly handle the keying speed required by the particular system and equipment. Time-division multiplex systems now in use on long-distance radiotelegraph circuits provide as many as eight separate channels over the one radio circuit.

10. **Checking the keying characteristics** of portions of the system, or of the entire system, is usually done by transmitting "reversals," a succession of dots of 50:50 mark-to-space ratio, by inserting in the automatic tape transmitter which keys the radio transmitter a closed loop of properly perforated tape. For speeds up to about 100 wpm the usual high-speed ink recorder can be used for checking character formation quite satisfactorily. For more accurate information, especially at higher speeds, the cathode-ray oscilloscope must be used. Amplifiers associated with the testing equipment must be better than the equipment being tested to avoid false deficiencies in the observation.

RADIOTELEGRAPHIC SERVICES

11. **Services.** Code-communication channels and equipment can be classified according to the type of service rendered by them under the general headings of point-to-point and mobile. The former is handled between stations on land while the latter is handled between shore and ship stations, or between ships.

12. **Long-distance point-to-point communications** are now generally handled on high frequencies. Frequencies used range from about 3,000 to 25,000 kc depending upon distance, season of the year, time of day, path traversed, etc. Proper choice of frequency allows reliable communication between any two points on the earth with transmitters of modern design. Power output of the equipment ranges from 1 to 50 kw. Owing to the extreme variations in transmission conditions encountered at these frequencies, it is necessary to have available at least 10 kw output from the transmitters for high-speed automatic operation over the longer distances. Even with the maximum output of present transmitters and with directive antennas for both transmission and reception, communication is slowed down or even stopped at times by severe disturbances in transmission conditions. By use of frequency-shift keying¹ (FSK), in which system signals are transmitted as marking and spacing signals on slightly different frequencies at substantially constant amplitude, considerably less transmitted power is required for a given circuit in comparison with that required for on-off keying. Recent tests have evaluated the power gain of FSK under certain conditions experienced in commercial service as 11 db, approximately.²

Normal field strengths obtained at the receiving antennas range from 0.1 up to 105 μv per m or more, depending on transmitter radiation, path, and transmission conditions. The minimum signal required for reliable commercial operation depends partly on the noise level at the receiving point. Atmospheric disturbances (static), while troublesome at times, are not so serious as in the case of low frequencies. Fading requires the use of a greater signal-to-noise ratio on high frequencies. Utilization of space, frequency, polarization, or time diversity of fading will overcome, to a great extent, the bad effects of static and permit successful operation on much weaker signals. A very rough estimate of the minimum field strength ordinarily required for code communication, with automatic recording, is 5 μv per m. Slow-speed aural reception can be carried on with field strengths of as low as 0.1 μv per m.

Minimum field strength required is determined by (1) directional distribution of *noise* at the receiving point, (2) directivity and pickup of the antenna system, which are both effective in determining the *gain* of the antenna, (3) the *noise equivalent* of the receiver itself.

Prior to 1928 long-distance communications were handled almost exclusively on frequencies ranging from about 15 to 30 kc. Great-circle distances covered on such commercial circuits ranged from 2,000 to 5,000 miles roughly. Such circuits are now

¹ Sprague, R. M., Frequency Shift Radiotelegraph and Teletype System, *Electronics*, November, 1944.

² Peterson, H. O., *et al.*, Observations and Comparisons on Radio Telegraph Signalling by Frequency Shift and On-Off Keying, *RCA Rev.*, March, 1946.

maintained by a few administrations for stand-by purposes, primarily to supplement h-f circuits when the latter are disturbed, but also for such special services as the dissemination of press and other multiple-addressed material to ships and other stations scattered over a considerable area.

Approximate values of field strength to be expected are calculated from the Austin-Cohen transmission formula:

$$E = 120\pi \frac{HI}{\lambda D} \sqrt{\frac{\theta}{\sin \theta}} \times e^{-u} \quad (2)$$

$$u = \frac{0.0014D}{\lambda^{0.6}}$$

where HI = effective height times current for transmitting antenna, m-amp

λ = wavelength, km

D = great-circle distance, km

θ = arc of great circle between transmitter and receiver

E = received field strength, μv per m

or the slightly different expression

$$E \text{ in } \frac{\mu V}{m} = \frac{377HI}{\lambda D} e^{-u} \quad (3)$$

where

$$u = \frac{0.005D}{\lambda^{1.25}}$$

which is derived from data taken on the New York to London circuits at frequencies ranging from 17 to 60 kc.¹

For successful operation on low frequencies, the received field strength required under average conditions is about 20 μv per m. It must be above the level of atmospheric disturbances and other local sources of noise to give readable signals. Automatic recording requires a signal to noise ratio of about 2:1 considering the general or average noise level. Moderately severe atmospheric disturbances such as crashes and clicks will be from several to ten times as strong as a normally satisfactory signal.

13. High Frequencies vs. Low Frequencies. Advantages of high frequencies for long-distance communication are (1) lower first cost of equipment and antennas, (2) smaller power consumption, (3) higher keying speeds, (4) less trouble from static, (5) directive transmission, (6) greater distances covered with a reasonable and practicable transmitter power. Disadvantages are (1) interruption of service due to severe magnetic disturbances, (2) effects of fading, (3) necessity for having several frequencies for each circuit for 24-hr service the year round.

Advantages of l-f operation are (1) freedom from interruption of service by magnetic disturbances and (2) comparative reliability and steadiness of signal strengths.

14. Medium-distance point-to-point communication by means of ground-wave transmission is used in the northern latitudes as an alternative to h-f sky-wave transmission which is frequently and severely disturbed by auroral phenomena in those regions. Frequencies in the order of 100 to 150 kc are satisfactory for circuits up to about 1,500 miles. In less disturbed areas frequencies of from 1,000 to 3,000 kc are used for nighttime sky-wave transmission, and from 3,000 to 8,000 kc for daytime transmission. Field strength requirements are of the same order as for long-distance communication either for automatic or aural reception.

15. Short-distance point-to-point communication may be effectively handled on frequencies extending into the v-h-f and u-h-f regions.² With frequencies over about 50 Mc relay stations (repeaters) are required where distances greater than line of sight must be covered. The usual distance between repeaters is from 20 to 40 miles depend-

¹ Espenschied, Lloyd, Anderson, C. N., and Bailey, Austin, Transatlantic Radio Telephone Transmission, *Proc. IRE*, February, 1926.

² Bramhall, F. B., Radio Relays for Telegraphy, *Elec. Eng.*, November, 1946.

ing upon the character of the terrain, the height of the antenna supports to be used, and the frequency. In commercial radiotelegraph service this type of transmission is mostly utilized for the remote-control links between the central office and the h-f radio stations. Because of the greater band width usable on the u-h-f circuits, it is customary to channelize the circuit into a group of voice-frequency bands, some of which are further subdivided into narrow-band channels for tone-telegraph service. The technique is very much the same as for carrier transmission on land lines. Such u-h-f circuits may be used to replace landline and cable circuits where installation and maintenance of the latter would be difficult. Since reception on these higher frequencies is remarkably free from atmospheric disturbance, the quality of these circuits equals or exceeds the quality of landline circuits.

16. Mobile communication is a different class of service in many respects from point-to-point. Except at the larger coastal stations and on a few ships, transmission is entirely by hand and copying is by ear. This is because of the nature of the traffic handled; a coast station usually has little more than 10 or 20 messages for any one ship at a time, and vice versa. Automatic transmission and reception are used only when traffic on hand amounts to some 40 messages or more. The same operator generally handles both transmission and reception, which is rarely the case in point-to-point service. Owing to the great number of ships served and to the intermittent nature of their traffic, the marine frequency bands must be shared by all ships. This creates interference and traffic-handling problems that are not encountered with point-to-point. For best service a marine operator must have direct control over the receiver tuning. Remote control is generally used for keying the transmitters at coastal stations, and often for changing frequencies, which permits separation of the transmitting station from the receiving station by a distance which often reaches 50 miles, so that simultaneous transmission and reception may be carried on without interference to reception.

The frequencies utilized mostly lie within the 350- to 500-kc band for short-distance work, the 500-kc frequency itself being reserved for calling and distress. A 5-kw coastal station has a normal daylight range of around 500 miles and a night range of several thousand miles under favorable conditions on this band. Ship-transmitter ranges are somewhat shorter unless working a shore station equipped with most efficient receivers and directional antennas. For long-distance communications the older large ships use frequencies on the order of 150 kc with an operating range under favorable conditions of from 1,500 to 3,000 miles. The newer ships rely upon high frequencies which provide ranges of the same or greater distance with comparatively low power.

Aeronautical radiotelegraph service is similar in most respects to marine service. It is limited in most cases now to use on long-distance overwater flights, radiotelephone being favored for short-distance work and for overland flights where communications may be established between the plane and the many airports encountered en route.

TRANSMITTING SYSTEMS AND EQUIPMENT

17. The Alexanderson alternator is used for l-f transoceanic code communication. This is a high-speed inductor type of machine having a large number of poles so that frequencies up to 30 kc and higher may be obtained directly. The alternator has an output of 200 kw and is driven by a 600-hp, two-phase induction motor through a set of gears to give the desired alternator speed. The stator has 64 separate windings connected to separate windings on the antenna input transformer. One winding is used to supply a tuned circuit, the output of which is rectified and used for automatic speed control. Such an alternator intended for operation at 27,200 cycles has 1,220 poles and is driven at a speed of 2,675 rpm.

To maintain the frequency constant to approximately 0.1 per cent and to have it the same under conditions of full load and practically no load, elaborate compensating means are provided as shown on the schematic diagram. Primary compensation saturation transformers each have an a-c and a d-c winding so connected that the volt-

age at the motor depends upon the impedance of these transformers which, in turn, depends upon the value of current in the d-c winding. Connected to the slip rings of the wound rotor are two banks of liquid rheostats, the "running" bank being connected at all times and the compensation bank being thrown on or off by the contactors. These contactors, and the contactor in the primary compensation d-c control circuit, are operated from a master relay which is controlled from the central traffic office. Compensation adjustments are made to maintain the machine at the same speed with the control key open or closed.

18. Method of Keying. Keying the output is accomplished by means of a magnetic modulator which is a special transformer having an a-c winding and a differentially connected d-c saturation winding. When the control key is open, a relay closes this d-c circuit, and the resulting drop in impedance of the a-c winding detunes the antenna

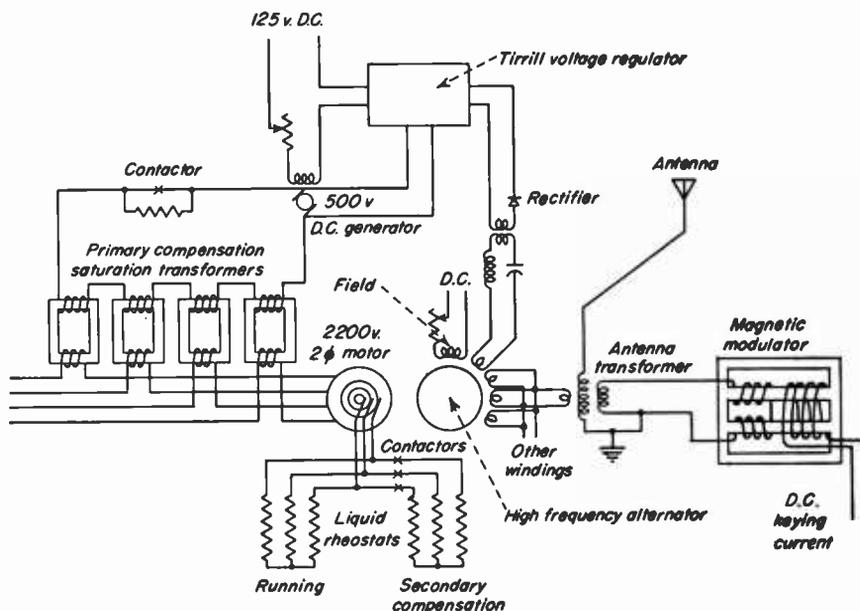


FIG. 3. Alexanderson alternator equipment.

and reduces the alternator output voltage so that practically no current circulates in the antenna circuit. For key closed, the d-c winding is deenergized and the antenna circuit becomes resonant to the alternator frequency, so that normal antenna current is obtained. Owing to the low frequency of the system and the low resistance of the antenna circuit, also on account of the large contactors required in the compensation circuits, keying speeds are limited to about 120 wpm.

19. Other means of generating 1-f waves, such as the Goldschmidt alternator, static frequency multipliers, and arc transmitters, have been used in the past but are now considered obsolete. The reader is referred to the several textbooks on radio communications published prior to 1940 and available in any technical library for description of this equipment.

Tube transmitters are now available for operation on frequencies as low as 14 kc since tubes to handle the enormous power required for long-distance communication on these frequencies have been developed to satisfy the demand for greater power on broadcast frequencies. Such transmitters, of the master-oscillator power-amplifier type, have been built with output ratings up to 500 kw.

20. L-f antennas of the various familiar types such as the T, inverted L, and umbrella have been used. Masts for these structures have, in some cases, been as high as 1,000 ft. Ordinarily they range from 400 to 800 ft in height. The technical problem is to get as many amperes in an antenna of as great an effective height as possible with a given power input. Voltages from antennas to ground may easily be 100 kv or more so that corona and insulation considerations place a limitation on the design. Of the total power supplied to the antenna, the useful portion is that radiated. The remainder is accounted for by conductor losses, coil losses, leakage, and corona (if present), and by loss in the resistance of the ground return path. In a structure where most of the capacitance is from the flattop to earth and where the dimensions are considerably less than a wavelength, the radiation resistance is given approximately by the relation $R = 1,600(H^2/\lambda^2)$, where H is the effective height of the antenna and λ the length of the radiated wave. Approximate calculation of H is possible in simple cases by summing up the products HI for all sections of the structure and dividing by the total current. This is done by calculating the capacitances to earth of the various sections, and by measuring the total value. Experimental methods of determining the capacitance from small-size models are described by Lindenblad and Brown.¹

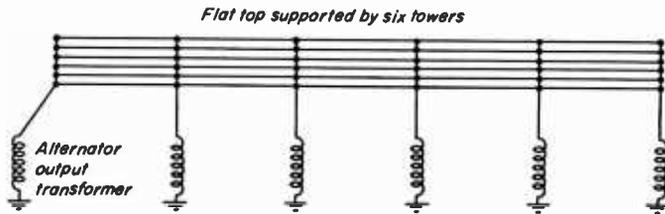


FIG. 4. Multiple-tuned antenna.

21. The multiple-tuned antenna consists of a long flattop supported by towers and having down-leads at a number of points which pass through tuning inductances to earth. The total antenna current is the sum of all the currents measured at the base of the tuning coils. A system of buried wires and overhead conductors connected to them through current-equalizing coils is laid out to give a uniform distribution of current in the earth under the antenna. This is approximately the condition for minimum earth resistance. This uniform distribution is sometimes altered, by experiment, to reduce the losses still further. Such antenna and ground systems often have a total resistance of less than $\frac{1}{2}$ ohm. Total antenna currents of 700 amp and more are obtained, by this means, from a transmitter output of 200 kw. For N tuning points the inductance of each down-lead and coil is approximately N times that which would resonate with the total antenna capacitance at the desired frequency. The physical length of such an antenna for operation at 17 kc, or thereabouts, may be 1 or $1\frac{1}{2}$ miles, with as many as six tuning points.

22. **Removal of Ice.** In climates where sleet is experienced the antenna wires should be counterweighted, rather than solidly anchored, in order to lessen the chances of breakage. A heavy coating of sleet on the wires, with the attendant increase in sag, throws the antenna out of tune as well as endangering it mechanically. When this becomes serious, it is necessary to melt the sleet from the wires in order to get normal antenna current. For this purpose, break insulators and by-pass capacitors are so arranged in the antenna wires that a series circuit of all (or part) of the wires is obtained at the low-power supply frequency. Special transformers supply power at about 2,000 volts for the purpose. This is sent through the antenna conductors just long enough to heat them sufficiently to melt off the sleet or ice.

¹ Lindenblad, N. E., and Brown, W. W., Main Considerations in Antenna Design, *Proc. IRE*, June, 1926.

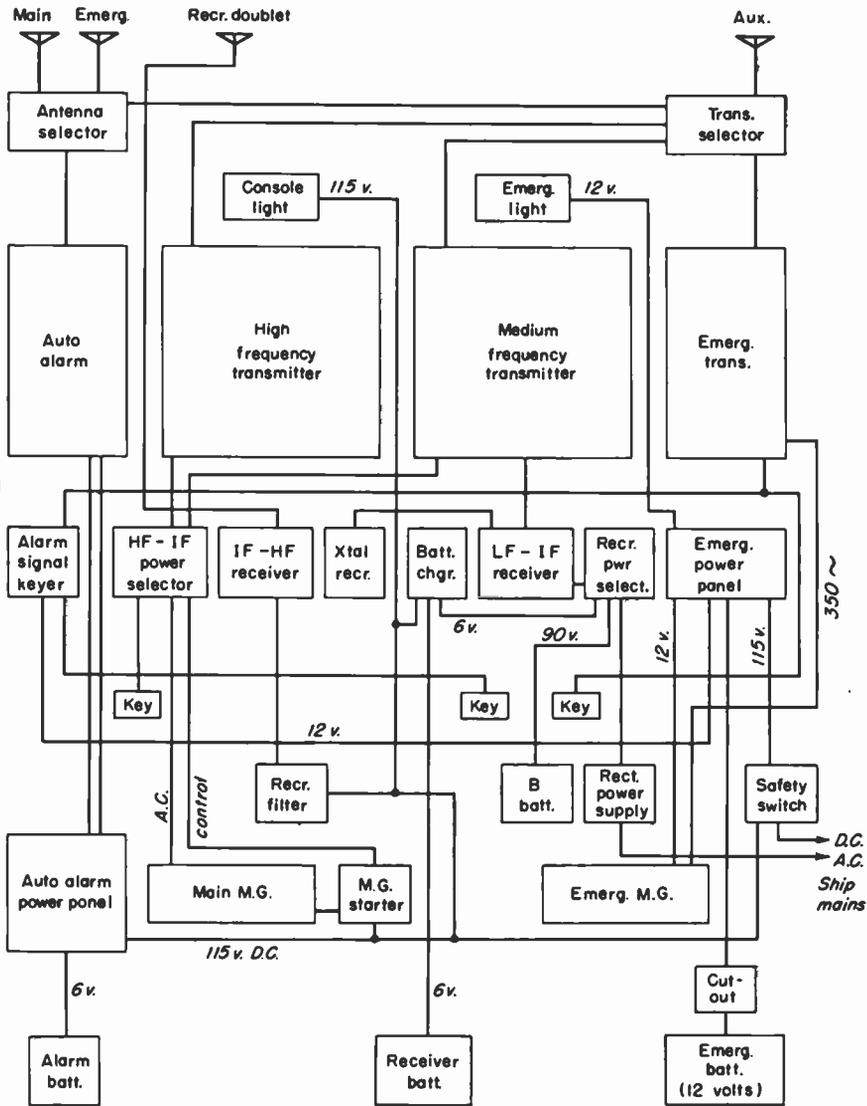


FIG. 5. Typical shipboard radio installation.

23. Marine Transmitters. Shore-station transmitters are little different in general from those used for point-to-point service. One convenient arrangement eminently suitable for the average station utilizes several r-f units, each complete from oscillator to 3-kw power amplifier with a single power-supply unit of ample size to permit simultaneous operation of the several r-f units if desired. A modulator unit may also be included to provide voice communication or tone telegraph. Simultaneous keying of the several r-f units will permit the listening vessel to determine which frequency

is best for its particular distance from the shore station and adjust its own transmitter to reply on the same frequency or one near to it. The more powerful stations use separate transmitters rated from 5 to 20 kw each. On the medium-frequency transmitters relays are usually provided in the r-f circuits to change the circuit constants by remote control so that one transmitter may serve two frequencies, say, 500 kc for calling and 425 kc for working. Other relays are provided to permit the transmitter to remain in an instantly available stand-by condition with filaments at reduced voltage and plate supply off for power conservation until keying starts, at which time proper voltages are applied.

24. Antennas used for marine shore stations are usually of the conventional T or L type supported on high towers (usually 300 ft) for medium frequencies, although some work has been done recently on applying the practice now common for broadcast transmission of using vertical tower radiators. For h-f transmission it is customary to use the simple doublet because of its nondirectional pattern, or a simple array to provide a limited amount of directivity seaward.

25. Transmitters for shipboard use are generally of lower power than those for coastal stations. Cost and space requirements are important factors which must be kept down. The usual equipment is, therefore, more simple and compact than that treated above. A type much in favor consists of a self-contained "console" or "package" assembly of the several units required for a complete shipboard radio station.¹ A typical assembly includes:

1. Main transmitter, 200 watts output, frequency range 350 to 500 kc, A1 and A2 emission
2. Emergency transmitter, 40 watts output, range 350-500 kc, A2 emission only
3. H-f transmitter, 200 watts, range 2,000 to 24,000 kc, A1 and A2 emission
4. Main and emergency receiver, range 15 to 650 kc
5. Auxiliary receiver, range 85 to 550, and 1,900 to 25,000 kc
6. Emergency crystal receiver, range 350 to 515 kc
7. Autoalarm for automatic reception of the international autoalarm signal on 500 kc
8. Alarm signal keyer for automatic transmission of the international autoalarm signal
9. Main motor generator operating on 115 volts d.c.
10. Emergency motor generator operating on 12-volt storage battery
11. Power switches and battery-charging panel
12. Antenna switching panel
13. Operating table with usual accessories such as lights, telegraph keys, and typewriter

Figure 5 is a block diagram of the console unit described above.

26. The h-f transmitter consists of an oscillator operating in the band of 1,500 to 3,000 kc followed by a first buffer, second buffer, and power amplifier. The buffers also act as frequency multipliers so that the fundamental, second, third, fourth, sixth, and eighth harmonics of the oscillator frequency may be obtained. With this arrangement only four crystals are required to provide one calling and one working frequency in each of the eight bands, as follows:

Crystal frequency, kc*	Output frequencies, kc				
	1	2	3	4	5
2,070 (c)	4,140	6,210	8,280	12,420	16,560
2,080 (w)	4,160	6,240	8,320	12,480	16,640
2,760 (c)	5,520	11,040	16,560	22,080	
2,765 (w)	5,630	11,060	16,590	22,120	

* (c) is a calling frequency; (w) is a working frequency.

Ganged switches for inductance taps and tuning capacitors permit quick selection and adjustment of any frequency desired. For keying the transmitter, a quick-acting

¹ McDonald and Hopkins, Marine Console Radio Unit, *Electronic Ind.*, January, 1944.

relay keys all stages in the transmitter. For A2 emission a modulation transformer which modulates the power-amplifier plate circuit obtains its input from a 500-cycle generator which is a part of the main motor-generator.

27. The medium-frequency transmitter consists of an oscillator and a power amplifier, both operating on the desired frequency. Eight pretuned frequencies are available: 355, 375, 400, 410, 425, 454, 468, and 500 kc. The eight oscillator inductors, one for each frequency, are each provided with an adjustable iron core which is locked at the proper position for frequency adjustment. The oscillator may be used as a self-excited master oscillator or as a crystal-controlled oscillator. The power amplifier uses a single tapped inductor for all eight frequencies. A modulator transformer is used to plate-modulate the power-amplifier stage for A2 emission, the secondary of this transformer being connected in series with the high-voltage d-c plate circuit of the power-amplifier tubes, the primary being connected to the 500-cycle part of the main motor-generator. All tubes in this set are quick heating thereby permitting the trans-

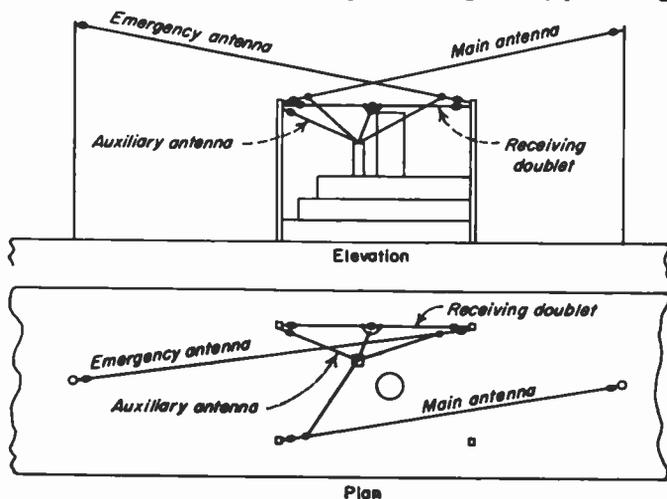


FIG. 6. Shipboard antenna arrangement (Victory class).

mitter to be keyed as soon as the motor-generator comes up to speed in about 5 sec. "Break-in" keying is provided by the keying relay so that the ship antenna is automatically connected to the receiver each time the transmitter key is opened.

The emergency transmitter consists of an oscillator and a power amplifier in a simplified low-power version of the main transmitter. All power is derived from a 12-volt storage battery which heats the filaments directly and drives a small motor-generator providing 350-cycle a.c. which is stepped up in voltage and applied to the plates of the power amplifier in a self-rectifying full-wave circuit. This method provides complete modulation with a 700-cycle note.

A typical arrangement of shipboard antennas is shown in Fig. 6 which is an installation as applied to the Victory class of cargo vessels. The main antenna and emergency antenna are each used for both medium-frequency and h-f transmission, as well as for medium-frequency reception.

28. Alarm Signal Keyer. This is an automatic device for operating the keying circuit of either the main or emergency transmitter to transmit repeatedly the international autoalarm signal which consists of a series of at least twelve 4-sec dashes, each separated from the other by a 1-sec space. The keyer consists of a vibrator operating on the 12-volt emergency battery and producing a 60-cycle voltage of about 22 volts which drives a synchronous motor. The motor gear train drives a timing cam at 12

rpm. The periphery of the cam is designed so that for four-fifths of its revolution it closes a switch to key the transmitter on the desired 4-sec dash.

29. H-f transmitters for point-to-point radiotelegraph service are characterized mainly by the use of class C amplification in the frequency-multiplying and output amplifier stages for maximum efficiency and power. These transmitters must maintain a high degree of frequency stability and be able to operate at high keying speeds. They generally feed a low-impedance transmission line for connection to a directive antenna.

Reduction of channel spacing resulting from the ever greater demand for frequency and channel assignments in the h-f range requires a corresponding greater stability of the frequency of the emitted carrier waves. To maintain a tolerance of better than ± 0.01 per cent requires the use of a very carefully stabilized and compensated tube oscillator or some control device such as a quartz crystal. Crystal control is in almost general use for this purpose and, even with this device, special means are used for obtaining the best results. Crystals used are either self-compensating for temperature variations or are enclosed together with their associated circuit elements in temperature-controlled compartments. Power supply to the crystal oscillator is maintained at constant voltage. Feedback and other detuning effects from following stages are prevented by careful isolation of the oscillator circuit.

Commercial h-f transmitters used for long-distance communication have an output of from 1 to 50 kw. The oscillator operates at a fraction of the final frequency desired, usually $\frac{1}{4}$ or $\frac{1}{8}$. This stage is followed by a buffer stage to isolate it from the later stages, then by one or more multiplying stages to the final amplifier stage at the output frequency. The lower power stages generally use tetrode- or pentode-type tubes and extremely careful shielding of the tubes and circuits including filtering of the power-supply leads to eliminate troublesome feedback effects without the use of neutralization. Except for the lower power types these transmitters use air- or water-cooled metal-anode power tubes in the final amplifier stage, and these must be provided with proper neutralization of the tube capacitances. The tank circuit of the power amplifier is coupled through a transmission line to the antenna. Means for harmonic suppression are usually provided in the coupling circuit.

Radiotelegraph transmitters in point-to-point service are usually operated on several frequencies over the 24-hr period. Frequency change is accomplished either by switching complete pretuned r-f chains of amplifiers, or by switching taps on inductances in the several stages combined with readjustment of variable capacitors in the tuned circuits.

For high-speed telegraphic operation using on-off keying, the voltage regulation of all plate and bias supplies must be good. If poor regulation exists, the envelope shape of the characters will be triangular or irregular instead of rectangular. A small amount of lag may be introduced intentionally to round off the corners to reduce the high-order side-band frequencies produced by the transients resulting from suddenly turning the r-f energy on and off, which side bands may produce severe interference in nearby receivers in the form of "key clicks." To provide good regulation hot-cathode mercury-vapor rectifiers are used for supplying the high d-c potentials required. These tubes, together with the high-voltage transformers used with them, have very good regulation at high values of output voltage.

30. Frequency-shift Keying. Transmitters used with the frequency-shift method of keying are not subject to the severe regulation requirement of on-off keyed transmitters. In this method of keying the r-f output remains substantially at full amplitude constantly, keying being accomplished by varying the instantaneous frequency between two limits: one being for the "mark" condition, the other for "space." In a frequency-shift method commonly used, the transmitter is keyed by a unit which is substituted for the usual oscillator stage. This unit consists of a crystal-controlled oscillator operating on a frequency which differs from the required r-f by 200 kc, and a stable 200-kc oscillator whose tuned-circuit frequency can be shifted by a predetermined amount, say, 106 cycles, upward or downward. The r-f oscillator is

modulated by the 200 kc plus 106 cycles for "mark" and 200 kc minus 106 cycles for "space" as determined by the keying control circuit. All modulation products except the sum of the crystal oscillator frequency and the 200 kc plus and minus 106 cycle oscillator frequency are filtered out, and the remaining r-f signal with minor side bands is fed into the first frequency-multiplying amplifier stage of the transmitter. The example given here is for a transmitter in which the output frequency is four times the oscillator frequency and the frequency shift between mark and space is about 850 cycles. The amount of frequency shift may be greater or less as required to meet the

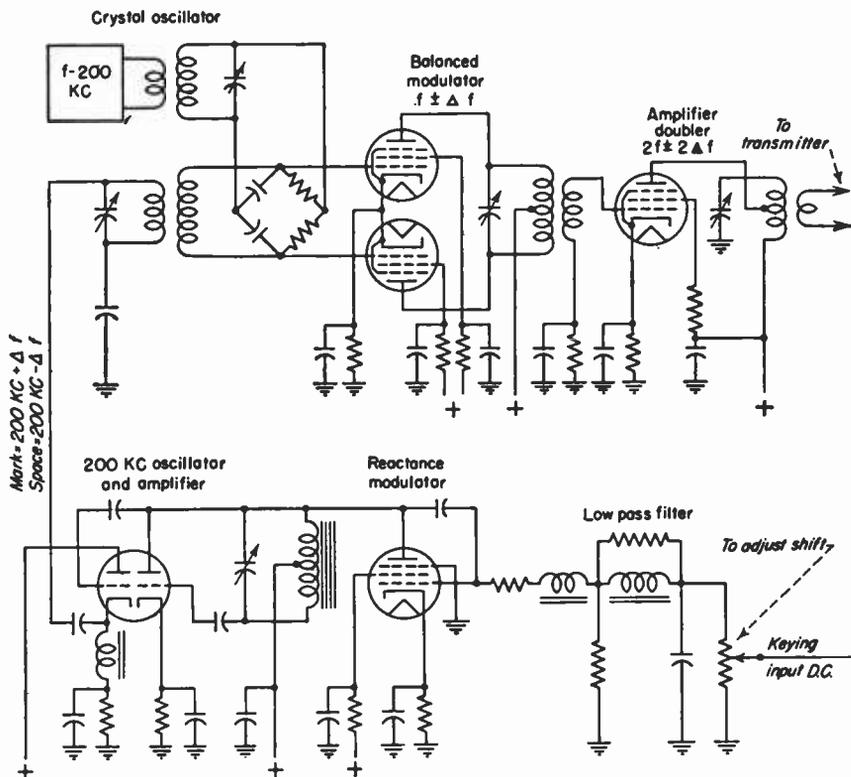


FIG. 7. Transmitter frequency-shift keyer, schematic.

particular service conditions involved. Figure 7 is a diagram of the FSK unit described.

31. Keying On-Off Signals. For continued operation at keying speeds greater than 100 wpm (80 cps) it is inadvisable to use a system of keying which employs electro-mechanical relays. A tube keying stage is therefore used to key one of the low-power stages of the transmitter. The FSK unit described above includes such a keying device. For on-off keying, several methods are available of which the one shown in Fig. 8 is particularly successful. Plate voltage to one of the transmitter low-power stages is fed in parallel to the plate of the keyer tube through a common resistance from the plate supply. When the key is down, a negative bias exceeding the cutoff value is placed on the keyer-tube grid so that this tube draws no current and allows normal voltage to be applied to the amplifier stage. When the key is up, the keyer-tube with a slightly positive voltage on the grid now draws plate current through the

series resistance with consequent reduction of voltage on the amplifier plate to a point where the amplifier does not sufficiently excite the next stage. All succeeding stages are also biased to keep from emitting without excitation from the previous stage in each case.

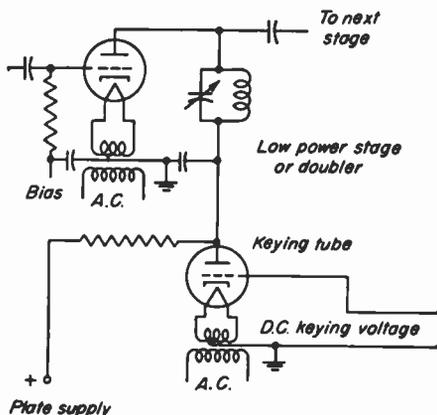


FIG. 8. Tube keyer for transmitter.

is usually preferred because of its ability to operate effectively over a relatively broad band of frequencies. This is an important economic consideration at a major transmitting station where many frequencies must be utilized to cover the varying conditions experienced over a period of years.

RECEIVING SYSTEMS AND EQUIPMENT

34. L-f Receivers. L-f receiving equipment must be designed to reduce trouble from static to a minimum and to separate transmitters differing in frequency by only about 200 cycles, which is the approximate spacing of assigned channels. The use of four efficient tuned circuits provides the required selectivity together with moderate ease of handling. For commercial work it has been the practice to obtain the h-f selectivity ahead of an aperiodic amplifier, then to go to a heterodyne detector of either the single-tube or balanced modulator type which is followed by as much a-f amplification as is required. The final selectivity may, if necessary, be obtained by the use of narrow band-pass a-f filters. For complete separation of signals on adjacent channels, this is often necessary. Owing to the difficulty of obtaining complete shielding of these comparatively low radio frequencies, it is generally advisable to use astatic pairs of coils in all tuned circuits, couplers, oscillators, etc., in addition to the use of a reasonable amount of shielding. Transformers and couplers are built with electrostatic shields to prevent capacitance coupling, where this is undesirable.

In a large receiving station where it may be necessary to receive a number of signals from approximately the same direction, a single aperiodic antenna system is the most economical and practical. The individual receivers are fed by means of "coupling tubes" operated from a common, or from individual antenna-output transformers. All tuning is done beyond these coupling tubes so that operation of each individual receiver is entirely independent of all others.

35. L-f Directional Antennas. Reduction of static is accomplished by the use of directive antenna systems. Arrays of large loops, or of loop and vertical combinations, are one means of obtaining directivity. Where the nature of the soil is such as to produce a considerable tilt of the wave front, the Beverage wave antenna¹ is used to advantage. This antenna consists of one or two wires strung on poles at a height of about 20 ft, and extending in the direction of the desired signal for a distance of

¹ Beverage, Rice, and Kellogg, *The Wave Antenna*, *Proc. AIEE*, March, 1923.

32. Modulation. For receiving systems that rely upon frequency diversity of fading, it is desirable to modulate the wave radiated from the transmitter at an a-f rate of something under 1,000 cycles. To prevent interference with signals on adjacent channels, this modulation should be reasonably free of harmonics. Otherwise the higher order of side bands will extend over into the adjacent channels.

33. H-f antennas used for point-to-point code transmission are in most cases directional arrays of dipoles or long wires in the several forms described in Chap. 14. Where ground space permits, the rhombic array of long wires

approximately 1 wavelength. The antenna is highly directional, and small signal voltages obtained from stations to the rear can be compensated out by feeding into the signal a small voltage of proper amplitude and phase obtained from the damping resistance connected between antenna and ground, or by setting up reflections in the antenna itself.

36. Marine receivers are usually provided in two forms: one to cover the l-f and medium-frequency bands from 15 to 650 kc, and the other to cover the h-f band up to about 25,000 kc. The h-f receiver may overlap the medium-frequency band as well.

A typical shipboard l-f receiver is of the tuned r-f type with an autodyne detector. Two stages of r-f amplification are provided with a gain control connected to the cathode circuits of both r-f tubes to increase the control-grid negative bias when the gain is to be reduced. The regeneration control varies the voltage impressed upon the screen of the detector. Resistor-capacitor networks couple the first and second audio tubes while an audio output transformer couples the second audio tube to a loud-speaker or headphones. Power supply is obtained from batteries so that reception will be assured even if the ship power fails. The equivalent shore-station receiver is somewhat the same but most often uses an a-c power supply in lieu of batteries.

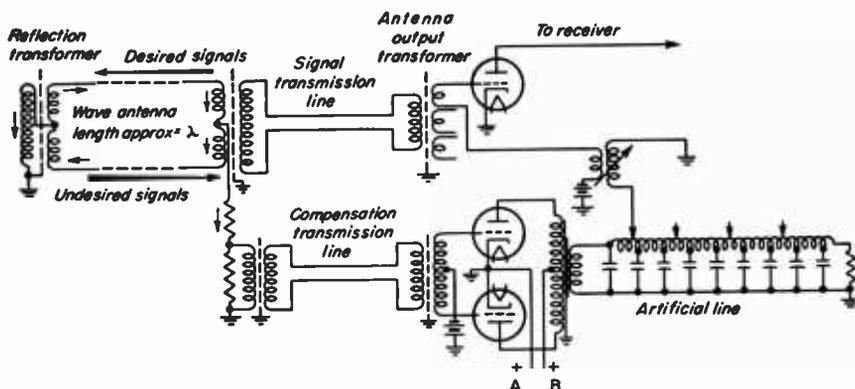


FIG. 9. Wave antenna and output circuits.

The associated h-f receiver is almost invariably of the superheterodyne type. Circuits used are conventional. The i.f. of a typical marine receiver is 1,700 kc. A v.c. and a beat-frequency oscillator (BFO) are provided. This receiver is designed to operate on 115 volts a.c. or d.c., hence it may be operated on the ship's power mains. Its frequency range covers 85 to 550 kc, as well as the h-f bands, hence this receiver may be used to monitor the busier medium-frequency channels in lieu of the other receiver and thereby conserve battery power.

In addition to the above, most vessels carry an emergency receiver, and many are provided with an automatic autoalarm. The emergency receiver is a simple crystal-detector type of receiver tunable over the range from 350 to 515 kc, by means of a fixed inductance and variable condenser.

37. The autoalarm is a device to stand watch on the 500-kc distress frequency when the radio operator is not on duty and receive the international autoalarm signal, which consists of a series of 4-sec dashes separated from each other by 1-sec spaces. Reception of this signal causes an alarm to sound on the ship's bridge indicating to the officer on watch that a distress signal has been picked up.

Basically the autoalarm is a superheterodyne broadly tuned to receive a signal whose frequency may extend from 487.5 to 512.5 kc. Its detector output controls a relay tube which has the coil of a relay connected in its plate circuit. This relay initiates the timing action of the selector circuit which follows. The selector performs three

functions: rejects signals having a duration less than 3.5 sec approximately; rejects signals having a duration greater than 4.5 sec approximately; and recognizes spaces between signals providing these spaces do not exceed 1.5 sec. This is accomplished by a series of three selector tubes with suitable *RC* networks to provide the proper time

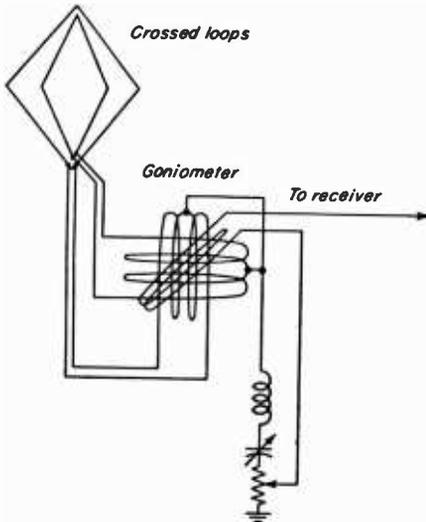


FIG. 10. Loop and vertical antenna for directive reception.

combination, giving a cardioid directive diagram, can be arranged with crossed loops and a goniometer so that the operator can rotate his antenna-reception diagram at will.

39. H-f receiving equipment for commercial point-to-point service must deliver a signal which is as nearly perfect as is possible. This requires a high degree of frequency stability, the best practicable over-all selectivity and means for reducing the effects of fading to a minimum. The receiver should have a total band width such that it will provide an attenuation of at least 60 db at the frequencies of the channels adjacent to that on which reception is being carried on. In calculating selectivity requirements, the assigned channel spacing must be reduced by twice the frequency tolerance permitted on each channel. This gives the frequency spacing between two signals on adjacent channels, when the frequencies of the two transmitters have drifted toward each other. Protection against all other types of interference, such as those encountered in superheterodyne receivers, should be not less than 70 db. At the same time, the useful band width must be sufficiently great so that no undue amount of attention will be required to keep signals fairly well centered in the pass band of the receiver. With present-day stability of transmitter frequencies and of receivers, this means a useful band width of from 1 to 4 kc depending upon the carrier frequency.

Present-day receivers,¹ to provide the required performance, are generally of the multiple-detection (superheterodyne) type in which one or two i-f systems are employed. It is only by the use of a relatively low final i.f. that the necessary selectivity and useful band widths can be obtained. The required i-f characteristics are obtained by use of either a band-pass filter or a number of stages of amplification employing one or more tuned transformers per stage. Choice of more than one band width in the i-f system is highly desirable and often necessary.

In equipment used for high-speed automatic operation, the signal is amplified, beat

¹ Moore, J. B., Recent Developments in Diversity Receiving Equipment, *RCA Res.*, July, 1937.

delays, and associated relays which serve to advance a stepping relay if the signal timing is correct or restore it to the starting point if the timing is incorrect. Four correctly timed dashes advance the stepping relay to a position where it causes a bell to ring to indicate reception of the distress signal.

38. Marine Receiving Antennas. At shore stations reduction of static and interference is accomplished by the use of directive antennas. For the lower frequency band the Beverage wave antenna has the advantage of relatively large pickup, good directivity with compensation, and the ability to supply a number of receivers operating at the same or different frequencies. Where reception from all directions is required and for the medium-frequency bands where the wave antenna is unsuitable for night reception, antennas of the flattop, inverted L, T, vertical, or loop types are employed. The loop and vertical

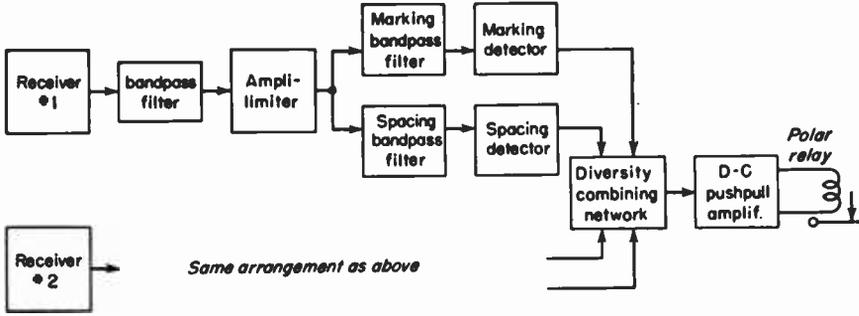
down to a lower frequency, and then rectified. The rectified output, consisting of short and long pulses of d.c., is used to operate a relay of the vacuum-tube type. The tube relay, or "keyer," controls the signal fed to the tone line from a local a-f source. The receiving operator is thus supplied with an audio signal of constant frequency and intensity regardless of any changes in the actual radio signal which are not great enough to make it drop out of the receiver. By means of a-f filters, six or more keyed tones of this sort may be handled over a single, two-wire tone line.

40. Diversity Reception. To minimize the effects of fading, receiving equipment is arranged to take advantage of the diversity of fading existing, at a given instant, either on slightly different frequencies at the same location or on the same frequency at points separated 10λ or more apart. Frequency diversity, in practice, is most economically obtained by modulating the carrier with an a.f. not higher than 1,000 cps, preferably not higher than 500 cps, to minimize interference to signals on adjacent channels. This results in radiation on the carrier and on an upper and a lower frequency. If the band width of the receiver is sufficient to pass these three frequencies and if the normal signal strength on any one of these frequencies is sufficient to operate the keying device, considerable diverse fading on the several frequencies received can be tolerated. In spite of the fact that a lesser peak voltage can be obtained from a modulated signal than from a pure c-w signal, considerable improvement is obtained, under practical conditions of fading, by its use. Where space diversity is utilized, a pure, unmodulated signal is preferred. In this case two or three separate receivers are fed from separate directive antennas spaced 10λ or more apart. The rectified outputs from these receivers are combined and made to operate the keying device. Confining the radiated energy to a single frequency means greater signal strength for a given transmitter power, and combination after rectification eliminates the consideration of instantaneous phase relations which might be such as to cancel rather than add.

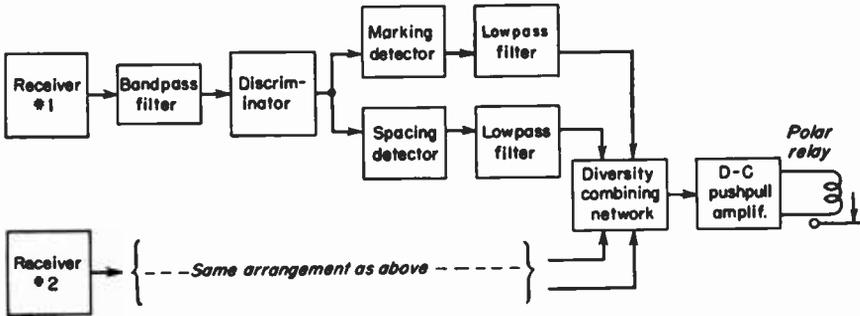
41. Limiting Circuits and Automatic Gain Control. Under conditions of high signal-noise ratio and violent fading, the use of considerable limiting in the receiving equipment is desirable. This should be done following the final selectivity and, for on-off keying, must be in a system having small enough time constants so that the decaying transients occurring after each overload do not occupy an appreciable portion of the interval between characters. To use such limiting successfully it is essential, as stated before, to pass up to about the fifth harmonic of the keying frequency. If this is not done, wide variations in mark/space ratio of the final signal will occur as the degree of limiting varies with the signal strength.

Recent practice has been to use some system of automatically controlling the gain (AGC) of the r-f amplifier stages. A particularly successful method operates as follows: A "threshold tube" (V_2 in Fig. 11) is provided to furnish a limiting bias to the r-f amplifier when no signal is being received. There is no bias on the grid of V_2 during no signal intervals so that anode current during this time is relatively high. Under this condition the bias applied to the grids of the r-f amplifiers is determined by the position of the arm on the threshold potentiometer P_2 . When an r-f signal is received, however, AGC potential for the r-f amplifiers is developed across the input potentiometer P_1 by the rectified signal current which flows through it to the diode output terminal of the i-f amplifier. For all signals above the adjustable threshold value the potential developed across P_1 is greater (more negative) than that across P_2 . Under this condition the potential across P_1 is the bias voltage applied to the r-f amplifiers. The automatic swinging back and forth from AGC bias supplied by the signal itself to the manually adjusted threshold bias is the chief feature of this circuit. It permits the use of AGC at practically all times, even when the signal-to-noise ratio is poor. It gives the desired AGC action on strong signals and prevents noise from breaking through during spaces or during pauses in transmission. The time constant of this system can be adjusted by connection of C_1 , C_2 , or C_3 .

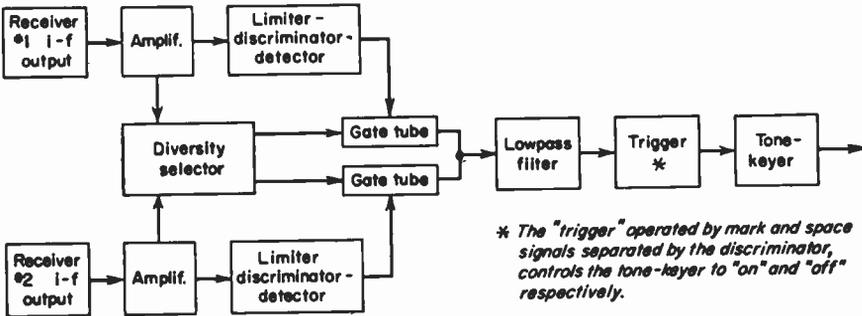
42. Frequency-shift keying reception may be accomplished by several methods of which the following are representative:



Audio frequency band-pass filter method



Audio frequency discriminator method



* The "trigger" operated by mark and space signals separated by the discriminator, controls the tone-keyer to "on" and "off" respectively.

I.F. discriminator method

FIG. 12. Frequency-shift keying receiving methods.

wire and two-wire types are used where the system is to be kept balanced with respect to earth. Antenna systems which operate against earth generally use the coaxial line in which the outer conductor is grounded. The two types of systems are sometimes connected together by means of suitably tuned transformers.

Antennas, particularly those with directional characteristics, must be so located as to guard against pickup of ignition and other man-made electrical noise as from passing automobiles and industrial activities. Receiving stations are consequently located in the open country, and even then "guard strips" several hundred feet wide are recommended for separating antennas from highways or residential districts.

44. H-f receiving antennas differ little in general from transmitting types commonly used, the major difference being the less stringent insulation requirement of the receiving antennas. The usual directive arrays of dipoles and long wires are equally satisfactory for reception and for transmission, those having the smallest rearward ears in their patterns being best for minimizing interference from transmitting stations on unfavorable bearings. Effective types peculiar to reception are the "fishbone antenna" and the "multiple unit steerable antenna" (MUSA).

45. Power supply for commercial receiving equipment must be absolutely reliable and not subject to interruption. Storage batteries operated on either a floating or a charge-and-discharge basis are used for this service at major stations.

Charging equipment consists of motor-generator sets for filament batteries, where relatively heavy currents are required, and either motor-generators or rectifiers for batteries of smaller rating such as used for plate and bias supply. Where receiving antennas may be located fairly close to the building that houses the charging equipment, this must be located in a specially shielded room to prevent direct radiation into the antennas. Equipment used for floating batteries that are in service must be provided with effective filtering between it and the battery and load bus.

Where the nature of the radio service does not warrant the expense of installing and maintaining storage batteries, reliance may have to be placed on the continuity and reliability of a-c power service provided by the local power company. In such cases the most economical and flexible arrangement for a small station is to provide each receiver with its own filament transformer and its own plate and bias supply rectifiers.

An emergency power supply should be provided for a-c operated equipment. Where storage batteries are installed for supplying the receivers during power failures, additional emergency power supply may or may not be necessary. In some cases the most economical arrangement may be a battery installation that will take care of only short-period outages and an emergency power plant to care for longer periods of failure of the public power service.

CONTROL METHODS AND EQUIPMENT

46. Central Office. In commercial radiotelegraphic systems the transmitters are controlled from a central traffic office¹ and received signals are conveyed to this central office from the receiving station by landlines. Transmitting and receiving stations are, in some cases, as much as 500 miles distant from the central office. The tendency, however, is to keep this distance below 100 miles to reduce initial and maintenance costs, or rentals, of landlines or radio control channels. Long control and tone lines are justified only if a distant location of the transmitter will effect a considerable saving in the power required to obtain satisfactory service, or if the distant receiving site is considerably superior to nearby ones in signal-noise ratio. In l-f transoceanic and medium-frequency marine work the use of long landlines is often well worth while. In h-f work the over-all results are not so dependent upon geographical location. Suitable sites are generally available within 100 miles of the city to be served.

47. Automatic Transmitters. In "automatic" operation of code circuits, a tough paper tape is perforated by means of a machine which has a keyboard similar to that of standard typewriters. This tape is then fed through the "automatic transmitter"

¹ Rau, D. S., and Brown, V. H., A Modern Radiotelegraph Control Center, *RCA Rev.*, July, 1939,

in which two cam-operated steel rods come up against the tape at every point where a perforation might exist. Where one is, the rod goes on through, and a contact operated by a lever on the lower end of the rod is closed. These two rods controlling the "make" and "break" contacts alternate in coming against the tape and are sufficiently offset in the direction of travel of the tape so that perforations in the upper (make) and lower (break) rows, when opposite the same center hole, give a dot and, when opposite adjacent center holes, give a dash.

The two contacts supply current in opposite directions to a polar relay which, in turn, keys the control circuit going to the transmitting station. For speeds above 100 wpm it is desirable to have as few mechanical relays as possible between the automatic transmitter and the keying circuit of the radio transmitter. The time required for a relay armature to travel from one contact to the other, although short, becomes important when the duration of a dot is less than 0.010 sec. Vacuum-tube keying devices may be used in place of mechanical relays for such high-speed service.

Printing-telegraph equipment employs a special model of automatic tape transmitter, which is adapted to the different code used for such systems.

In installations of multiplex equipment employing the principle of time division, automatic tape transmitters supplying the several channels are usually synchronized and phased to give the required over-all performance of the multiplex system.

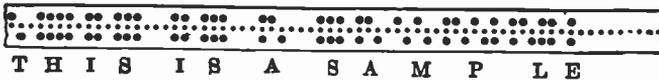


FIG. 13. Sample tape of automatic transmitter.

48. Tone-control Circuits. Where only a few transmitters are to be controlled from one point, d-c polar keying is the most economical and satisfactory. A complete metallic circuit is to be preferred to a single wire with ground return, although the latter is entirely satisfactory in many cases.

In a large central-office system the number of control lines required can be greatly reduced by the use of multiple tone, or "voice-frequency carrier" control. By the use of a number of different frequencies and band-pass filters at both ends of the circuit, as many as 12 channels can be obtained on a voice-frequency channel which will pass frequencies from about 400 up to 2,500 cycles with approximately equal attenuation. In one such type of equipment, the a-f supply is a multifrequency inductor-type alternator having a separate winding and rotor for each frequency. Energy from this machine is keyed by means of either electromechanical or vacuum-tube relays which are controlled by the automatic tape transmitter. Band-pass filters in the individual control channels reduce the harmonic content of the signal supplied to the line to a low value and also round off the corners of the square keying envelopes.

The band width required in filters for tone-control work depends upon (1) the maximum keying speed to be handled and (2) the fidelity of envelope shape required for the particular application. Where great fidelity is not required or where the over-all transmission gain of line and associated equipment does not vary more than about 20 per cent, it is sufficient to pass the second harmonic of the keying frequency. This means a total band width of four times the keying frequency. To obtain fairly square envelope shape, with a mark-to-space ratio of about 60:40, it is necessary to pass up to the third harmonic or a total band of six times the keying frequency, at least.

For the lengths of line normally used between central offices and outlying stations and for present-day code keying speeds, the matter of phase distortion due to the line is of relatively small importance.

49. Control equipment used at transmitting stations may be of either the d-c or tone-operated type, depending upon the system used at the central office. In a double-current d-c system the conventional polarized telegraph relay is used as a main-line relay for speeds up to some hundred words per minute. Where normal operating speeds run much above 100 wpm special high-speed relays of the polarized type must

be used. Large keying and compensation relays and contactors used in l-f transmitters are controlled by the line relay or a heavier intermediate relay. In tube sets, especially h-f equipment, higher keying speeds are possible and require the use of a minimum number of mechanical relays. For d-c control the main-line relay may operate directly into a tube keyer incorporated in the transmitter.

In tone-control systems the equipment at the transmitting station comprises band-pass filters and amplifier-rectifier units. The rectified output may be used to operate either electromechanical relays or tube keyers. Where such equipment is used at large high-powered transmitting stations, it may have to be protected from stray fields of the transmitters, transmission lines, and antennas. The amount and disposition of shielding and filtering required by control equipment and associated wiring depend on numerous factors such as the following: (1) lowest frequency radio transmitter installed or contemplated, (2) highest control frequency planned, (3) r-f field intensities, and (4) level of control signals and voltages. It is obvious that a high-power l-f transmitter operating on a frequency of about 20 kc will create serious problems where it is desired to employ control channels ranging in frequency from, say, 400 to 20,000 cps or higher.

Tube keyers, while more elaborate than the usual mechanical relays, are capable of operating at practically any speed desired. They also eliminate relay maintenance and adjustment. In the simpler arrangements, the control tone is amplified, rectified by either a diode or a triode rectifier, then passed through a smoothing circuit or low-pass filter. The d-c pulses obtained are applied to the control elements of the keying-stage tube or tubes.

50. Received Signal Transfer. Systems for transferring signals from the receiving station to the central office are similar to the transmitter-control systems. In h-f work the actual radio signal, after heterodyne detection, is amplified and rectified and applied to a tube keyer. This may be arranged to supply d.c., or tone, for transfer to the traffic office. A-f filters, of the same type used for tone control, allow a number of channels to be handled over one line.

Where tone lines are long enough to require the use of one or more repeaters, care must be taken that the sum of the voltages of all channels is not so high as to cause any overloading of the repeaters. If this takes place, intermodulation between channels will be caused, which results in mutilated signals at the central office. With lines having repeaters and the usual band-pass filters, it is essential that all channels be kept at approximately the same signal level. A maximum difference of 2:1 between any two channels should not be exceeded. Large differences in channel levels are apt to cause interference on the weaker ones.

In m-f and l-f receiving stations the contacts of all telegraph keys and relays must be prevented from sparking, and the wires to and from the contacts must be properly filtered. If these precautions are not taken, serious click interference will be experienced in the receiving equipment. The same applies to commutator-type electric motors. Circuit breakers should preferably be located in a shielded room.

TRANSCRIBING METHODS AND EQUIPMENT

51. High-speed Reception. As the average operator copies at a rate of only about 40 wpm, aural reception must be replaced by some method in which a record is made of the signal on the high-speed circuits, the recorded signal then being copied off at a slower speed by one or more operators. The older dictaphone and photographic methods of recording were not entirely satisfactory. Most systems which utilize the Morse code now use some form of "ink recorder" in which the movement of a pen is controlled by the incoming signal and makes short and long characters on a moving paper tape.

Reception by tape has the double advantage of speed and of providing a record to which the operator may refer in case any question arises.

52. Ink Recorder. One commonly used type of ink recorder consists of a small coil suspended in a strong unidirectional magnetic field supplied by an electromagnet.

The signal is amplified and rectified and the d-c pulses sent through the recorder coil which, in turn, moves the pen arm up against an upper stop. With no signal current flowing, the pen is held against the lower stop by the spring of the pen arm and coil suspension. To improve the action of the device at high speeds, the coil is suspended midway between the stops, and current reversals are used in place of pulsating d.c. to operate the coil. This is obtained from a pole-changing relay operated by the re-

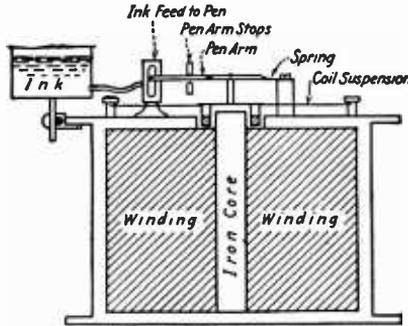
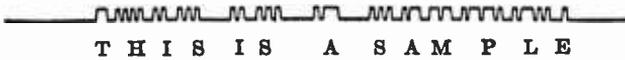


FIG. 14. Ink recorder. Paper tape and tape guide not shown.

tified signal, or from a special amplifier-rectifier unit which gives an output d.c. in opposite directions for "mark" and "space."

53. Printers. Where printing-telegraph equipment is employed, manual transcription of the incoming signal is eliminated. The printed tape coming from a tape printer is simply pasted on message blanks. Errors may be corrected by obtaining the required correction from the distant radio terminal and pasting it over the original which contained the error. On circuits of high quality where a minimum of errors



occur, reception may be on a "page printer" directly on a form ready for delivery to the addressee without further handling. At large central offices, printer reception is now made on a "reperforator," which produces a tape perforated with the code characters corresponding to the incoming signal. This tape is then transferred to one of a group of automatic transmitters for further transmission of the message to another destination, or to the delivery desk of the same office, at which point the message is then printed upon the message blank. This system of transferring messages at a central office from one circuit to another, commonly referred to as the "semiautomatic method," is currently in favor with the major radiotelegraph companies.¹

¹ Sparks, S., and Kreer, R. G., Tape Relay System for Radiotelegraph Operation, *RCA Rev.*, September, 1947.

INDEX

A

- Absorption, dielectric, 169-170
- Absorption modulators, 482-483
- A-c impedance relations, 200-201
- A.c.-d.c. power supply, 311-312
- Acoustical coefficients, 764
- Acoustics, room (*see* Room acoustics)
- ADF, 1137
- Admittance, 71-80
- A-f range for broadcasting, 896-900
- Air-borne installations, 1151-1156
- Air navigation (*see* Radio aids to navigation)
- Air traffic control, 1071, 1144-1146
- Aircraft antennas, 633-634
- Aircraft receivers, 884-885
- Alarm, automatic signal, 1171
- Alarm receivers, automatic, 883-884, 1175
- Alexanderson alternator, 1166
- Alloys, magnetic, 148-149
 - permanent-magnet (*see* Permanent-magnet alloys)
- Alnico, 150
- Alternator, Alexanderson, 1166
 - Goldschmidt, 1167
- Altimeters, 890-891, 1135-1136
- Aluminum-backed screens, 600-601
- AM (*see* Amplitude modulation)
- Amateur receivers, 877-878
- Ammeters, a-c characteristics of, 262
- Ampère, 41
 - circuitual law of, 47-48
- Amplification, a-f, per stage, 347-348
 - noise limitations of, 1020
 - r-f, grid input capacitance effect in, 392
 - video (*see* Video amplification)
- Amplifiers, audio (*see* Audio amplifiers)
 - audio-frequency, 346-389
 - cathode-coupled, 1019
 - cathode follower, 377-378, 403
 - d-c, 363-364
 - differentiator, 442-443, 1029
 - direct-coupled, 363-364
 - Doherty high-efficiency, 413-414, 951
 - driver, 489-490
 - feedback, 373-378
 - grounded-grid, 401-403
 - i-f, 813-816
 - impedance-capacitance-coupled, 352-354, 393-394
 - limiting, 939
 - Loftin-White, 363-364
 - modulated, 412-414
 - multistage, 349-364
 - overdriven, 442
 - phase-inverter, 362-363
 - power (*see* Power amplifiers)
 - power supply for, 313, 360-362
 - push-pull, 362, 369-373
 - RC (*see* RC amplifiers)
- Amplifiers, r-f, 390-415, 797-809
 - band-pass filters in, 397-399
 - cascade, 396-399
 - class B, 411
 - class C, 408
 - impedance-coupled, 393-394
 - instability of, 400
 - LC coupling of, 396
 - linear, 949-951
 - power, 406
 - regeneration of, 399-401
 - tuning stability of, 399
 - use of ceramic capacitors with, 399
- speech-input, 921
- transformer-coupled (*see* Transformer-coupled amplifier)
- video, transient response of, 1019-1020
- wide-band, 1024
- Amplitude characteristics of electrical communication system, 23-25
- Amplitude control of oscillation, 431
- Amplitude limiters, 519-520
- Amplitude modulation, 461, 462-464, 482-488
 - vs. frequency modulation, 964-966
 - in nonlinear circuits, 479-488
- Amplitude-modulation detectors, 500-512
- Antennas, 602-683
 - aircraft, 633-634
 - Beverage, 620-623
 - broadcast, 623
 - current distribution in, 605-609
 - d-f, 1108
 - dipole, 634-645
 - directive, stability of, 632
 - directive broadcast, 627
 - dummy, 286
 - fishbone, 653-654
 - folded dipole, 641-642
 - h-f, 634-655
 - insulation for, 617-619
 - lazy-H, 639-640
 - long wire, 645-647
 - loop, 792-797
 - low-frequency, 609-610
 - master, 893
 - measurements of, 613, 673, 983-985
 - multiple-tuned, 1168
 - quadrant, 642
 - radio range, 619
 - reactance, 613-614
 - rhombic, 648-653
 - scale-model, measurements of, 613
 - ship, 633
 - terminology for, 603-604
 - u-h-f, 655-677
 - vertical dipole, 640-641
 - v-h-f, 655-677
 - measurements of, 673
 - whip, 619-620
- Aperture distortion, 1052
- Applegate diagram, 454

- Arc light, zirconium, 1058
 Arc transmitters, 1167
 Armstrong f-m method, 964
 Arrays, 627-632
 bedspring or billboard, 664
 dipole, 642-645
 Articulation, 732
 Aspect ratio, 996
 Atmospheric effects on propagation, 534-536
 Atom model, 26
 Atomic particles, 27
 Attenuators, fixed, in wave guides, 714-715
 variable, 715-716
 for vu meter, 271
 Audio amplifiers, 346-389
 equalization for, 379
 gain for, 378-379
 measurements for, 380
 Audio-frequency lines, 228-229
 Audio generator, electrostatic, 455
 Audio oscillator, Scott, 432
 Audio-transformer design, 355-357
 Autoalarm, 1175
 Automatic volume control, 813-820
 Automobile receivers, 856-857
 Ayerton-Mather universal shunt, 259
- B**
- Band-pass coupled circuits, 220-221
 Band-pass filters in r-f amplifiers, 397-399
 Band-pass r-f circuits, 223-224
 Band-spread circuits, design of, 832
 Band widths, and carrier frequency for radio
 service, 3, 5
 and gain equations, 221-223
 Bank windings, 127
 Bar, 9
 Barkhausen-Kurtz oscillator, 449-450, 574
 Battery chargers, 318-319
 Beacons, marker, 1078-1079, 1129
 Beam modulation tubes, 587-596
 Beam traveling-wave tubes, 596
 Beam-tube power amplifiers, 365-367, 579-580
 Beat-frequency oscillators, 436-438
 Bel, 8, 15
 Benito, 1120
 Beverage antenna, 620-623
 Boella effect, 97
 Bonding, and shielding, 884
 of tower radiators, 626-627
 Breakup, cone, 751
 Bridge, Carey Foster, 292, 305-306
 d-c measurements of, 291-307
 guard circuit of, 300-301
 Hay, 302
 Heaviside equal-arm, 305
 Kelvin, 293
 Maxwell, 302
 power supply and null detector for, 294-295
 rat-race wave-guide, 713
 resistance balance of, 298-299
 resonance, 303
 Schering, 299
 series Owen, 302
 slide-wire, 292-293
 stabilization of, 429-431
 Thomas's, 299
 Wheatstone, 291
 Wien, 304
 Bridge errors, 296-298
 Bridge structures, wave-guide, 711-713
- Bridge transformers, 295-296
 Broadcasting, 895-995
 a-f range in, 896-900
 international, 949
 radio facilities in, 938-945
 station coverage in, 986-992
 v-h-f, 992
 volume range in, 900
 (See also Radiobroadcasting)
 Broadcasting antennas, 623
 directive, 627
 Butterfly tuners, 861-862
- C**
- Cable circuits, equalizer for, 230
 submarine, 1064
 Cadmium cell, 255
 Camera, signal for, 1003
 television, 1006
 Capacitance, 58-62, 163-187
 calculation of, 170
 definition of, 163
 distributed effects of, on inductance, 119
 effect of frequency on, 171
 grid input, effect of, on r-f amplification, 392
 and inductance, comparison of, 301-303
 interelectrode, 348, 582-584
 measurement of, 123, 184-187, 275-276, 299
 Capacitance standards, 183-184, 255
 Capacitance units, 163
 Capacitors, ceramic, in r-f amplifiers, 399
 color code for, 172
 dielectric, energy stored in, 64
 electrolytic, 175-177, 344
 filter (see Filter capacitors)
 gang, 178-179
 inductive, 171
 noninductive, 171
 paper, 345
 ratings of, 344
 variable, design of, 177-183
 Capacity, specific inductive, 164
 Carbon microphones, 909-910
 Carey Foster bridge, 292, 305-306
 Carrier suppression, 463
 Carrier wave, 5
 Cascade amplifiers, r-f, 396-399
 Cathode, oxide-coated, 549, 552
 photocathode, 557, 1008
 preparation of, 551-552
 virtual, 566
 Cathode-coupled amplifier, 1019
 Cathode follower, 377, 403-404
 Cathode-ray-tube meters, 282-283
 Cathode-ray tubes, 599-601
 timing axis for, 284
 Cavity resonators, 718-730
 shunt impedance of, 720
 Cavity wavemeters, 729-730
 Ceramic capacitors in r-f amplifiers, 399
 Charge density, 30-31
 Choke, filter, design of, 337-344
 swinging, 337
 Choke coil windings, impregnation of, 343-344
 Chromaticity diagram, 1040
 Circuit, concepts of, 55-56
 coupled, 217-223
 damped, 70
 d-c, restoration of, 1032
 Eccles-Jordan, 277
 electric, 56-73

- Circuit, linear, 57, 58
 magnetic, 73-75
 neutralizing, 404-406, 949
 noise limiting, 821-823
 oscillatory, 70
 for out-of-phase voltages and currents, 225
 printed, 116
 pulsed, 193-198
 Q, in *LCR*, 198-205
 squelch, 821-823
 tapped tank, 215-216
 voltage-doubling, 317-318
 Circuit design, parallel-resonant, 211-213
 Circuit elements, lumped, 57
 voltage, current, and energy relations in, 68
 Circuit parameters, lumped, 54
 Circuit solutions, 68
 Clippers, diode, 521
 Coaxial line, 239
 Coaxial-line resonators, 723-725
 Code, color (*see* Color code)
 radio, 1159-1161
 Code printers, 1183
 Code receiving systems, 1174-1180
 Code speeds, 1163
 Code transcribing methods, 1182-1183
 Coefficients, acoustical, 764
 Coercive force, 146
 Coils (*see* Inductors)
 Color code, capacitance, 172
 chassis, 892
 loud-speaker, 893
 resistor, 101-102
 transformer, 330, 892
 Colpitts oscillator, 418
 Communication facilities, CAA, 1074, 1082-1083
 Comol, 150
 Comparison measurements, 286-291
 Compensation, h-f, 1015
 series-peaking, 1017-1018
 shunt-peaking, 1016
 Compensation theorem, 65-66
 Condenser microphone, 908-909
 Condenser speaker, 756
 Conductance, 58-60, 71
 variational, 60
 Conduction, electrical, 44-45
 metallic, theory of, 45
 Conductivity, 45
 Conductor, 28
 Conductor losses in wave guides, 694
 Conductor materials, 85
 Conductor temperature coefficient of resistivity,
 45
 Cone breakup, 751
 Consol, 1103-1104
 Contact potentials, 565-567
 Control resistors, concentric dual, 112-114
 miniature, 112
 Converters, crystal, 499-500, 806-807
 and mixers, 497-500
 tubes for, 580-582
 Copper-oxide rectifiers, 107, 267, 320, 480
 Copper-sulfide rectifiers, 320
 Core, powdered-iron, 141-142
 Core loss, 325-326
 tests for, 153-158
 Coulomb's law, 31-32, 39, 46
 Coupled circuits, 217-223
 Couplers, directional, 717-718
 Coupling, to cavities, 725-728
 coefficient of, 217
 Cross modulation, 792, 798
 Crystal converters, 499-500, 806-807
 Crystal cuts, piezoelectric, 421-424
 Crystal diode modulators, 488
 Crystal microphone, 907-908
 Crystal oscillators, piezoelectric, 419-426
 mountings for, 425-426
 temperature control of, 424-425
 Crystal speaker, 756
 Cunico, 150
 Cunife, 150
 Curl, 38
 Current, definition of, 254
 density of, 43
 distribution of, in antennas, 605-609
 electric, 42
 high-frequency meters for, 263-266
 measuring instruments for, 256-272
 out-of-phase, circuits for, 225
 sheet radiator, 661-662
 space-charge-limited, 563
 transient, 188-198
 (*See also* Voltage)
 Cyclophon tubes, 871
- D
- Damped circuit, 70
 Damping and period of instruments, 257-258
 Daraf, 62
 D-c measurements (*see* Measurements, d-c)
 Decca, 1117-1118
 Decibel, 8, 15
 table of, 273
 Decrement, logarithmic, 193, 257
 Deflection indicators, 257
 Delay networks, 231-233
 Density, charge, 30-31
 current, 43
 Detection and modulation, 461-522
 Detectors, a-m, 500-512
 balanced, 513-515
 diode, 500-505
 f-m, 512-519, 846
 grid, 505-506
 harmonic distortion in, 518-519
 heterodyne, 509
 high-level plate, 506-508
 infinite-impedance, 508-509
 phase, 519
 p-m, 512-519
 ratio, 515-517, 846-850
 square-law, 506
 standing-wave, 713-718
 video, 1026-1027
 Dielectric, 28
 breakdown voltage of, 61
 energy stored in, 64
 relative permittivity of, 61
 Dielectric absorption, 169-170
 Dielectric constant, 164-165
 Dielectric materials, 164
 table of, 167-168
 Dielectric polarization, 36-37
 Dielectric strength, 169
 Dielectric wave guides, 697-699
 Differentiator amplifiers, 442-443, 1029
 Dimensions and units, 78-82
 Diode clippers, 521
 Diode design, 563
 Diode detectors, 500-505

- Dipole, 634-645
 - electric, 655-659
 - magnetic, 659-661
 - Dipole arrays, 642-645
 - Direct-coupled amplifiers, 363-364
 - Direction finders, 882-883, 1080-1081, 1105-1108
 - Discriminators, 824-825
 - frequency, 225-226
 - Disk recording, 924, 929
 - Distance-measuring equipment (DME), 1077, 1118-1121
 - Distortion, intermodulation, 853-854
 - transmission without, 25-26
 - Divergence, 38
 - Diversity reception, 1177
 - Doherty high-efficiency amplifier, 413-414, 951
 - Dummy antenna, 286
 - Dynamic limiters, 850
 - noise-suppression, 837
 - Dynatron oscillator, 433-434
- E
- Ear characteristics, 851
 - Earth characteristics, 525
 - Eccles-Jordan circuit, 277
 - Echo boxes, 730
 - Eddy-current loss, 119
 - Einthoven galvanometer, 260-261
 - Elastance, 60, 61
 - Electric charges, 29
 - Electric currents, 42
 - magnetic field produced by, 43
 - Electric displacement, 35-36
 - Electric field, 29-30
 - Electric field intensity, 31
 - Electric flux, 35
 - Electric flux density, 35-36
 - Electrical communication principles, 2-7
 - Electrical conduction, 44-45
 - Electrical measurements, 254-307
 - Electrical sheet and strip, 148
 - Electrical units, international, 254
 - Electrodynamics, 41-51
 - Electrodynamometer, 261-263
 - Electrolytic capacitors, 175-177, 344
 - Electromagnetic field, 51-56
 - and electric circuit concepts, 55-56
 - energy in, 76-77
 - Electromagnetic spectrum, 1-2
 - Electromagnetic waves, 75-78
 - Electromagnetism, 41
 - fundamentals of, 26-28
 - Electrometer, Kelvin absolute, 273
 - Electromotive force, 48, 50
 - Electron, properties of, 27-28
 - Electron beam, deflection of, 1033
 - velocity of, 1033
 - Electron emission, field, 561-563
 - grid, 562
 - secondary, 559-561
 - Electron-tube meters, 279-286
 - Electron tubes (*see* Tubes)
 - Electrophoresis, 552
 - Electrostatic voltmeters, 273
 - Electrostatics, 28-38
 - potential, 33-34
 - Emf, induced, Faraday's circuital law for, 48-49
 - Emission, 545-563
 - efficiency of, 551
 - field, 561-563
 - photoelectric, 555-558
 - Emission, secondary, 559-561
 - testing of, 554
 - Enclosure, vented, 758, 760
 - Ensi, 782
 - Epstein core loss test, 153-154
 - Equalizers, 201, 207, 229, 230
 - Equipotential surface, 33
 - Equivalence of T and π networks, 247
 - Exponential representation of harmonic functions, 19-21

F

 - Facsimile, 1051-1067
 - propagation effect on, 1063-1065
 - Facsimile modulation, 1055
 - subcarrier f-m, 1061
 - Facsimile recording, direct, 1058-1059
 - mechanisms for, 1060
 - variable-width, 1058
 - Facsimile reproduction, accuracy of, 1051
 - Facsimile scanning, 1052-1056
 - Facsimile standards, 1066
 - Facsimile synchronizing, 1062-1063
 - Facsimile tape system, 1065
 - Faders, 918-919
 - Fading, 536-539
 - selective, 536
 - Faraday, 41
 - circuital law of, for induced emf, 48-49
 - Feedback, degenerative, 953
 - Feedback amplifiers, 373-378
 - Feedback oscillators, 417-419
 - Felici balance, 123-124, 305
 - Field emission, 561-563
 - Field intensity, magnetic, 40
 - measurements of, 987-988
 - Fields, interlaced, 997
 - Filament, base, 553
 - carbured, design of, 549
 - flat or ribbon, 551
 - tungsten, 546-547
 - thoriated, 547
 - Film engraving, 923
 - Filter capacitor ratings, 344
 - Filter capacitors, 335-336
 - electrolytic, 344
 - paper, 345
 - Filter rectifiers, 332-344
 - Filtering and shielding, 827-828
 - Filters, band-elimination, 250
 - band-pass, 251
 - in r-f amplifiers, 397-399
 - choke design of, 337-344
 - decoupling, 224-225
 - design formulas for, 249-253
 - high-pass, 250
 - low-pass, 250
 - RC, 233, 334
 - scratch, 201, 207
 - wave, 246-253
 - Fishbone antennas, 653-654
 - Flicker frequency, 14
 - Flutter, 926
 - Flux, electric, 35
 - magnetic (*see* Magnetic flux)
 - Flux density, electric, 35-36
 - Flux linkage, 49
 - Force, coercive, 146
 - electromotive, 48, 50
 - lines of, 30
 - magnetizing, 145

Force, magnetomotive, 47, 73, 144
 tubes of, 30
 Form factor, 309
 Fourier analysis, 21-22
 Fourier series, 17
 Frequency, comparison of, 287-288
 effect of, on capacitance, 171
 maximum usable, 528
 and pitch, 10-11
 and power range for communication, 9-10
 stability of, 255-256, 419
 television, allocation of, 1023
 transition, 834
 Frequency characteristics of electrical communication system, 23-25
 Frequency control, automatic, 824-827
 Frequency conversion, 461, 479
 Frequency converters and mixers, 497-500
 Frequency deviation, 466, 956
 Frequency discriminator, 225-226
 Frequency measurement, 276-278
 Frequency meter, 276-277
 Frequency modulation, 466-470, 496-497
 vs. amplitude modulation, 964-966
 Armstrong, 964
 carrier stabilization for, 957
 modulators for, 495-497
 Frequency-modulation detectors, 512-519, 846
 Frequency-modulation receivers, 842-852
 Frequency-modulation systems, 956
 Frequency monitor, 942
 Frequency multipliers, 414-415, 521, 1167
 Frequency-shift keying, 1172, 1177-1179

G

Gain control, automatic, 1177
 Galvanometers, a-c, characteristics of, table of, 261
 d-c, characteristics of, 258
 differential, 259
 Einthoven string, 260-261
 vibration, 260
 Gang capacitors, 178-179
 Gauss's theorem, 37, 40, 52
 Gee, 1114-1115
 Generators, electrostatic audio, 455
 impulse, 440
 photoaudio, 455
 saw-tooth, 1034-1035
 shading correction, 1007
 standard-signal, 285-286
 synchronization signal, 1010
 video-signal, 1005-1006
 (See also Oscillators)
 Gill-Moreil oscillator, 449
 Glide path, 1127-1129
 Glow-lamp recorder, 924
 Goldschmidt alternator, 1167
 Grid detectors, 505-506
 Grid emission, 562
 Grid input capacitance, effect of, on r-f amplification, 392
 Grid modulators, 483
 Ground-controlled approach (GCA), 1080, 1131-1133
 Ground systems, 615-617, 627
 Grounded-grid circuit, 859-861
 Guard circuit bridge, 300-301

H

Hammond brake speed control, 1057
 Harmonic distortion in f-m detector, 518-519

Harmonic functions, simple, 17-21
 Harmonic oscillators, 416
 Harmonic suppression, 970-974
 Hartley law, 3
 Hartley oscillator, 418
 design of, 458-459
 Hay bridge, 302
 Hazeltine Lanac system, 1146-1148
 Hearing, characteristics of, 737-740
 threshold of, 9
 Heaviside equal-arm bridge, 305
 Heising constant-current modulator, 954
 Henry, 118
 Heterodyne detectors, 509
 H-f antennas, 634-655
 H-f compensation, 1015-1016
 H-f current meters, 263-266
 H-f oscillators, 445-455
 H-f resistance, 86
 H-f resistors, characteristics of, 97-98
 composition of, 103
 Hipernic, 149
 Homing beacons, 1079
 Honeycomb coils, 126
 Horns, 746-749
 Humidity characteristics, 100
 Hyperbolic navigation systems, 1090-1093
 Hysteresis, 119, 146-147

I

Ice removal, 1168
 Iconoscope, 1006-1007
 IFF, 881
 Ignitron, 319
 Image dissector, 1009
 Image orthicon, 1008-1009
 Impedance, 69-71
 comparison of, 288-289
 equivalent, 201-203
 input, 348-349
 measurement of, 275
 mutual radiation, 741-744
 radiation, 740
 series, shunt loss in, 201-203
 shunt, of cavity resonators, 720
 of wave guides, 692
 Impedance-capacitance-coupled amplifier, 352-354, 393-394
 Impedance-matching transformers, 367-369, 716
 Impedance meters, 713-718
 Impedance relations, a-c, 200-201
 in r-f lines, 241
 Impedance transformers, 716
 Impregnation, of choke coil windings, 343-344
 of transformer windings, 343-344
 Impulse generators, 440
 Incremental permeability, 74, 146
 at low inductance, 155-158
 Indicators, d-f, 1109-1110
 tuning, 820-821, 846
 Inductance, 58-64, 118
 calculation of, 132-139
 and capacitances, comparison of, 301-303
 distributed capacitance effects of, 119
 measurements of, 120-124
 of iron-core coils, 159-160
 at low frequencies, 159
 mutual, 123-124, 305
 Turner constant-impedance, 160-161
 two-frequency test on, 122-123
 Inductance matching, apparent, 121

- Induction, 325
 intrinsic, 145
 low, incremental permeability at, 155-158
 magnetic, 40, 145
 residual, 146
 Induction regulator, 286-287
 Inductive tuning, 797-798
 Inductivity, 164
 Inductors, bank windings, 127
 finish, 140-141
 honeycomb coils, 126
 iron-cored (*see* Iron-cored inductors)
 magnetic field of, energy stored in, 65
 multilayer solenoid, 125
 permeability-tuned, 128
 powdered-iron, 130
 progressive universal windings, 127-128
 Q of, 119-120
 single-layer solenoid, 124-125
 toroidal iron-core coils, 129
 types of, 124
 universal windings, 125-126
 Input impedance, 348-349
 Input transformer design, 371
 Instruments, period and damping of, 257-258
 voltage-measuring, 272-274
 Insulation, for antennas, 617-619
 wire, 139-140
 Insulator, 28
 Integrator circuit, 1029
 Intensity, magnetic, Ampère's circuital law for, 47-48
 Interelectrode capacitance, effects of, 348
 Interference, 470-479, 1182
 from a-m stations, 472
 from f-m stations, 472-473
 Interference ratios, 992
 Intermodulation, 479
 Interrogation systems, 1139-1142
 Intrinsic induction, 145
 Ion spot, 1034
 Ion trap gun, 600
 Ionizers, metal-bristle, 1155
 Ionosphere, 526-530, 539-541
 Iron-cored inductors, 128-129
 inductance of, measurement of, 159-160
 toroidal, 129
- K**
- Kelvin absolute electrometer, 273
 Kelvin bridge, 293
 Kennelly-Heaviside layers, 526
 Kerr cell, 923, 1058
 Keying, frequency-shift, 1172, 1177-1179
 of on-off signals, 1173
 Kirchhoff's laws, 65
 Klystron efficiency, 590
 Klystron oscillators, 452-455
 Klystrons, 587-592
 reflex, 863
 two-cavity, 590-592
- L**
- Lanac system, Hazeltine, 1146-1148
 Landing system (ILS), 1079, 1123-1135
 Laplace equation, 37-38, 40
 LC table, 206
 LCR circuit transients, 191-193
 LCR circuits, *Q* in, 198-205
 Lens's law, 49
- Level indicators, 269
 standard reference, 900-901
 L-f compensation, 1018-1019
 Light-valve recorder, 923
 Lighthouse triode, 860
 Lighting, obstruction, 626
 Limiters, 440-442
 amplitude, 519-520
 dynamic, 850
 noise-suppression, 837
 threshold, 521
 Lines, a-f, 228-229
 artificial, 230
 coaxial, 239
 of force, 30
 pulse-shaping, 443
 r-f, general properties of, 235-246
 impedance relations in, 241
 stub, 245
 table of, 240
 voltage and power ratings of, 245-246
 six-wire, 239, 241
 (*See also* Transmission lines)
 Lissajous figures, 283
 Lobe switching, 665-666
 Loftin-White amplifier, 363-364
 Logarithmic decrement, 193, 257
 Logarithmic measurements, 15-16
 Loop antennas, 792-797
 Loran, 891, 1077-1078, 1110-1113
 Loud-speakers, 731-774
 color code, 893
 complete, 756-762
 condenser, 756
 crystal, 756
 diaphragms of, 749-752
 electrodynamic, 772
 high-fidelity, 854-855
 magnet for, 161
 motor for, 752
 moving-coil, 754
 tests for, 768-773
 transmission-line, 761-762
 Loudness level, 9, 737
 LR circuit transients, 190
- M**
- Magnetic alloys, 148-149
 Magnetic circuits, 73-75
 Magnetic dipoles, 659-661
 Magnetic field of inductor, energy in, 65
 Magnetic field intensity, 40
 produced by electric currents, 43
 Magnetic flux, 40-41, 49, 144
 divergence of, 43-44
 Magnetic flux density, 40-41, 49
 Magnetic induction, 40, 145
 Magnetic intensity, Ampère's circuital law of, 47-48
 Magnetic materials, 118, 143-162
 applications of, 161
 Magnetic permeability, 74, 146
 Magnetic permeance, 144
 Magnetic poles, Coulomb's law for, 39
 Magnetic recording, 161, 925-938
 Magnetic reluctance, 144
 Magnetic testing, 151-161
 Magnetism, nature of, 145-146
 Magnetizing force, 145
 Magnetomotive force, 47, 73, 144
 Magnetostatics, 38-41

- Magnetostriction, 256
 Magnetostriction oscillators, 426-427
 Magnetron oscillators, 451-452
 Magnetrons, 592-596
 adjustment of, 594-595
 interdigital, 595-596
 modes of, 594
 Magnets, loud-speaker, 161
 permanent, 161, 256, 755
 for meters, 256-257
 tests for, 158-159
 relay, 161
 Marine receivers, 877
 Marine transmitters, 1169
 Marker beacons, 1078-1079, 1129
 Maxwell bridge, 302
 Maxwell equations, 28, 51, 75
 Meacham oscillator, 429
 Measurements, 247-307
 antenna, 613, 673, 983-985
 scale-model, 613
 v-h-f, 673
 audio-amplifier, 380
 bridge, d-c, 291-307
 capacitance (*see* Capacitance, measurement of)
 comparison, 286-291
 d-c, 308-309
 indicating instruments for, 308-309
 direct capacitance, 299-300
 frequency, 276-278
 impedance, 275
 of inductance (*see* Inductance, measurements of)
 noise, 542-543
 power-factor, 276
 three-voltmeter method for, 288
 of pulsating currents and potentials, 272
 resistance (*see* Resistance measurements)
 Measuring instruments, voltage, 272-274
 Mercury-arc rectifiers, 319
 Mercury-vapor rectifier, 318
 Meters, cathode-ray-tube, 282-283
 electron-tube, 279-286
 frequency, 276-277
 high-frequency current, 263-266
 impedance, 713-718
 moving-diaphragm, 278-279
 rectifier, 266-272
 thermocouple, 263-266
 vacuum-tube, 279-286
 vibrating-reed, 277
 vu, attenuators for, table of, 271
 wavelength, 277
 Metric units, 79
 Micro-H, 1121-1122
 Microphone calibration, 913-914
 Microphones, carbon, 909-910
 cardioid, 905
 condenser, 908-909
 crystal, 907-908
 dynamic, 903-904
 parabolic, 910-913
 placement of, 914-922
 polydirectional, 906-907
 requirements for, 901-902
 testing of, 913-914
 unidirectional ribbon, 904-905
 velocity, 902-903
 Microwave ILS, 1133-1135
 Mixers and converters, 497-500
 MKS system, 82
 Modulated amplifiers, 412-414
 Modulation, 842
 amplitude (*see* Amplitude modulation)
 cross, 792, 798
 and detection, 461-522
 effect of Q on, 212
 equipment for, 954
 frequency (*see* Frequency modulation)
 monitor, 942-945
 in nonlinear circuits, 479-488
 phase, 466-470, 956
 pulse, 461-466, 488, 495
 pulse-code, 872-873
 pulse-time, 870-872
 velocity, 461
 video, 1021-1023
 Modulators absorption, 482-483
 amplitude, 482-488
 balanced, 485
 copper-oxide, 487-488
 crystal diode, 488
 facsimile, 1055
 frequency, 496-497
 f-m, 495-497
 grid, 483
 Heising constant-current, 954
 plate, 485-487
 p-m, 495-497
 pulse, 488-495
 reactance-tube, 957
 subcarrier f-m, 1061
 video, 1021-1023
 Mosaic, 1008
 Moving-diaphragm meters, 278-279
 Moving iron instruments, 263
 Mu factor, 571-572
 Mu-slop of tube, 572
 Multiplex, 1163
 Multiplier phototube, 557-559
 Multipliers, frequency (*see* Frequency multipliers)
 Multivibrator, 438
 one-shot, 491
 Mumetal, 149
 Music characteristics, 85, 733-737
 Musical instruments, peak power of, 900
 Mutual-inductance balance, Felici, 305
- N
- Navaglobe, 1104-1105
 Navar, 1142
 Navigational systems, accuracy of, 1122-1123
 circular, 1118-1123
 Gee, 1114-1115
 hyperbolic, 1110
 radial, 1093-1110
 radio, 1087
 Negative-resistance oscillators, 433-435
 Neper, 8, 15
 Networks, delay, 231-233
 d-f, 1081
 differentiating and integrating, 195-198
 dividing, 854
 electric, 86
 π , equivalence of, 247
 pulse-forming, 491-493
 T, 306-307
 equivalence of, 247
 and transmission lines, 226
 (*See also* Lines; Transmission lines)
 Neutralising circuits, 404-406, 949
 Neutrodyne, 404-405
 Neutron, 26

- Newtons, 46
 Nicol prism, 1058
 Noise, 473-479, 541
 atmospheric, 543-544
 characteristics of, 737
 fluctuation, 788
 in f-m receivers, 475-476
 impulse, 471, 476-479
 man-made, 544
 random, 3, 471
 in receivers, 786-792
 suppression of, 837, 857
 thermal-agitation, 788
 Noise levels, 739
 Noise-limiting circuits, 821-823
 Noise measurements, 542-543
 Norton's theorem, 66
 Null detector for bridge, 294-295
- O
- Oersted, 41
 Ohm's law, 65
 Ohmic resistance, 60
 Ohmmeter, direct-reading, 274
 Operator $j = \sqrt{-1}$, 19
 Optical lever, 257
 Orthacoustic system, 935
 Orthicon, image, 1008-1009
 Orthiconoscope, 1007-1008
 Oscillation, amplitude control of, 431
 parasitic, 967-970
 Oscillators, 416-460
 automatic-frequency-control, 455-456
 Barkhausen-Kurtz, 449-450, 574
 beat-frequency, 436-438
 blocking, 444, 493-494
 circuits for, 70
 classification of, 416-417
 Colpitts, 418
 dynatron, 433-434
 electron-coupled, 456-457
 feedback, 417-419
 Gill-Morell, 449
 grounded-grid, 450
 harmonic, 416
 Hartley, 418
 design of, 458-459
 high-frequency, 445-455
 klystron, 452-455
 low-frequency, 436
 magnetostriction, 426-427
 magnetron, 451-452
 Meacham, 429
 mechanical-electronic, 455
 modulated, 487
 negative-resistance, 433-435
 phase-shift, 432
 piezoelectric, 419-426
 (See also Crystal oscillators)
 power, design of, 457-459
 pulsed, 488-495
 RC, 431-433
 relaxation, 416, 438-440
 resistance-capacitance, 431-433
 resistance-stabilization, 427-428
 resonant line, 446
 ring, 448
 Scott audio, 432
 semiconductor, 435-436
 synchronized, 520-521
 timing, 444
- Oscillators, tracking problems of, 122, 213-215, 808-809
 transitron, 434-435
 tuned filter, 427
 tuning-fork, 455
 ultraaudion, 445
 vacuum-tube, 416-460
 Wien-bridge, 430
 Owen bridge, series, 302
- P
- Pain, threshold of, 9
 Panoramic receivers, 886-887
 Parabolic microphone, 910-913
 Parallel resonance, 207-213
 Parallel-resonant circuit design, 211-213
 Parameters, lumped circuit, 54
 Peak power of musical instruments, 900
 Pedestal, 1002
 Pentode tubes, 576-579
 Period and damping of instruments, 257-258
 Periodic functions, 22-23
 Permalloy, 149
 Permanent-magnet alloys, 140-151
 for meters, 256-257
 tests of, 158-159
 uses of, 161
 (See also Magnets, permanent)
 Permeability, 41, 47
 incremental, 74, 146
 of inductors, 128
 at low inductions, 155-158
 magnetic, 74, 146
 Permeability tuning, 399, 808
 Permeance, magnetic, 144
 in series and parallel magnetic circuits, 74
 Permendur, 149
 Perminvar, 149
 Permittivity, 47, 60
 of typical dielectrics, 61
 Perveance, 567
 Phase characteristics of electrical communication system, 23-25
 Phase detectors, 519
 Phase difference, 165
 Phase inverter, 362-363
 Phase modulation, 466-470, 956
 Phase-modulation waves, detectors for, 512-519
 modulators for, 495-497
 Phase-shift discriminator, 514
 Phase-shift effect on human response, 11
 Phase-shift oscillator, 432
 Phase velocity, 688
 Phasitron, 497
 Phon, 16
 Phonograph attachment, wireless, 836
 Phonograph-radio combination, 832-838
 Phosphors, 601, 1031
 Photoaudio generator, 455
 Photocathode, 557, 1008
 Photoelectric emission, 555-558
 Phototubes, 555-559
 multiplier, 557-559
 Physiological characteristics of sound, 8-9
 Physiological responses and communication system, 7-16
 Picture tubes, 1030
 Piezoelectric crystal cuts, 421-424
 (See also Crystals)
 Piezoelectric crystal mountings, 425-426
 Piezoelectric crystal oscillators, 419-426

- Piezoelectric crystals, temperature control of, 424-425
- Piezoelectric effect, 279
- Piston directivity, 744-746
- Pitch and frequency, 10-11
- Plate detectors, high-level, 506-508
- Plate modulators, 485-487
- Plate resistance, 572-573
- PM (*see* Phase modulation)
- Poisson equation, 37-38, 40
- Polarization, circular, 603
- dielectric, 36-37
- Polyphase rectifiers, 313, 316
- POPI, 1118
- Potential difference, 50
- Potential gradient, 34-35
- Potentials, contact, 565-567
- and pulsating current measurements, 272
- Powdered-iron cores, 141-142
- Power, dissipation of, in resistor, 64
- peak, of musical instruments, 900
- Power amplifiers, 364-373, 406
- beam-tube, 365-367, 579-580
- class B, 372-373
- push-pull, 369-373
- r-f, 406-415
- Power-carrying capacity of wave guides, 692-693
- Power factor, 120, 165
- measurement of, 276, 288
- Power level instruments, 269-272
- Power loss, 165-166
- Power-oscillator design, 457-459
- Power range and frequency for communication, 9-10
- Power ratings for r-f lines, 245-246
- Power-supply systems, 308-345, 966
- a.c.-d.c., 311-312
- audio amplifier, 313, 360-362
- null detector for bridge, 294-295
- ripple, 334
- television receiver, 309-310, 1033
- Power transfer, maximum, conditions for, 228
- Power transfer theorem, 66
- Power transformer, iron, 161
- Poynting's vector, 77
- Precipitation static, 885, 1154-1156
- Preemphasis, 844-845, 965
- Printed circuits, 116, 842
- Program monitor, 939
- recording, 922-938
- Progressive universal winding, 127-128
- Propagation, radio-wave, 523-544
- atmospheric effects on, 534-536
- beyond line of sight, 530-536
- effect of, on facsimile, 1063-1065
- line-of-sight, 533-534
- of sky wave, 526-530
- of surface wave, 524, 526
- u-h-f, 530
- Protons, 26
- Pulsating current and potentials, measurements of, 272
- Pulse-code modulation, 872-873
- Pulse formation, 490-491
- Pulse modulation, 461-466, 488, 495
- Pulse-shaping lines, 443
- Pulse-time modulation, 870-872
- Pulse-type communications, receivers for, 870-873
- Pulsed circuits, 193-198
- Pulses, synchronizing, 1004-1005
- Purkinje effect, 13
- Push-button tuning, 809-812
- Push-pull power amplification, 362, 369-373
- Q
- Q, effect of, on modulation, 212
- of inductance, 119-120
- in LCR circuits, 198-205
- measurement of, 123, 211
- representative values of, 200
- resonator values of, 719-720
- Q meter, 205
- R
- Radar, 889-890
- pulse-altimeter, 1136
- Radar range, multiple-track, 1115
- Radial navigation systems, 1087
- Radiation, conical scanning, 667
- Radiation impedance, 740
- Radiation pattern, vertical, 625
- Radiation resistance, 610-613
- Radiators, current sheet, 661-662
- parabolic reflector, 673-674
- tower, bonding of, 626-627
- Radin, 877
- Radio aids to navigation, 1068-1158
- air traffic control, 1071
- civil aviation facilities, 1072-1074
- collision warning, 1071
- final approach, 1070
- functional requirements of, 1070
- landing, 1070
- long-distance, 1017
- radio stations for, 1075-1077
- r-f bands, 1085
- short-distance, 1070
- Radio beacons, 881
- Radiobroadcasting, 895-995
- facilities for, 938-945
- secondary coverage in, 528
- transmitters for, 945
- Radio codes, 1159-1161
- Radio communication, basis of, 1-82
- Radio-navigation systems, 1087-1158
- accuracy of, 1122-1123
- Radio-phonograph combinations, 832-838
- Radio range, omnidirectional, 1096-1101
- Radio range antennas, 619
- Radio range beacon operation, 1094
- Radio-wave propagation (*see* Propagation, radio-wave)
- Radio waves, characteristics of, 523-524
- Railroad receivers, 875-877
- Ratio detector, 515-517, 846-850
- RC amplifiers, 349-352, 390-393
- charts of, 382-389
- compensation in, 226
- RC circuit transients, 189
- RC filter, 233
- Reactance, 70-71, 255
- of antenna, 613-614
- Reactance stabilization, 428-429
- Reactance-tube modulator, 957
- Reactance tubes, 825
- Reactance-variation method of measuring resistance, 290
- Receivers, a.c.-d.c., 840-842
- aircraft, 884-885
- amateur, 877-878
- automatic alarm, 883-884
- automobile, 856-857

- Receiving center, commercial, 1178
 Receiving systems, 775-894
 code, 1174-1180
 countermeasure, 881-882
 diversity, 887-888
 f-m, 842-852
 frequency-shift keying, 1172, 1177-1179
 high-fidelity, 851-856
 loran, 891
 marine, 877
 military, 879-882
 mobile, 873-875
 multiband, 831-832
 naval, 879-882
 noise, 786-792
 in f-m receivers, 475-476
 impulse, 478-479
 panoramic, 886-887
 for pulse-type communications, 870-873
 railroad, 875-877
 reflex, 842
 single-signal, 878
 superheterodyne, 498-499, 776, 799-809
 superregenerator, 776, 863-870
 testing, 777-786
 tracking, 122, 808-809
 underwriters' requirements for, 891-892
 v-h-f, 857-863
 wide-band, 888-889
 Reciprocity theorem, 66
 Recorder, glow-lamp, 924
 home, 837
 light-valve, 923
 magnetic wire and tape, 935-938
 Recording, constant-amplitude, 833-834
 constant-velocity, 833
 disk, 924, 929
 facsimile (*see* Facsimile recording)
 film-engraving, 923
 ink, 1182
 Kerr cell, 923, 1058
 magnetic, 161, 925-938
 orthacoustic, 935
 photographic, 1058
 program, 922-938
 re-recording, 931
 variable-area, 922
 variable-density, 922, 1058
 vibrating-ribbon, 923
 wax preparation for, 930-931
 Rectifier meters, 266-272
 Rectifiers, carborundum, 107
 copper-oxide, 107, 267, 320, 480
 copper-sulfide, 320
 dry metallic, circuits for, 323
 filters for, 332-344
 galena, 107
 lead-sulphide, 107
 mercury-arc, 319
 mercury-vapor, 318
 polyphase, 313, 316
 selenium, 107, 321-322
 vacuum-type, 314-316
 Rectigon tubes, 318
 Reflecting screens, flat, 671-673
 Reflectors, parabolic, radiators with, 673-674
 Reflex klystron, 863
 Reflex receivers, 842
 Relaxation oscillators, 416, 438-440
 Relay magnets, 161
 Reluctance, 73, 144
 in series and parallel magnetic circuits, 74
 Reluctivity, 41, 74
 Remote control and telemetering, 828-831
 Remote receiver, 811
 Re-recording, 931
 Residual induction, 146
 Resistance, 58, 70, 83-117
 critical damping, 258
 definition of, 255
 h-f, 86
 nonlinear, 59
 ohmic, 60
 plate, 572-573
 radiation, 610-613
 specific, 84
 temperature coefficient of, 84-85
 variational, 60
 Resistance balance of bridge, 298-299
 Resistance-capacitance oscillators, 431-433
 Resistance-coupled amplifier charts, 382-389
 Resistance curve, volume-control, 115
 Resistance measurements, 274
 reactance-variation method for, 290
 resistance-variation method for, 289-290
 resonance-rise method for, 291
 susceptance-variation method for, 290-291
 Resistance pads, 233-235
 Resistance stabilization of oscillators, 427-428
 Resistance values, preferred, 103
 Resistance-variation method for measurement, 289-290
 Resistivity, 45, 59, 84
 and temperature coefficient, 45
 Resistors, adjustable, 93
 carbon, 96
 ceramic, 94
 color code, 101-102
 composition, 103-105, 111-114
 control (*see* Control resistors)
 deposited carbon, 106-107
 filament-coated, 96
 fixed composition-type, 95-97
 forced air-cooled, 104
 humidity characteristics of, 100
 immersion of, in oil, 104
 metal-clad, 94
 negative temperature coefficient of, 110-111
 noise characteristic of, 98-100
 operation of, at fractional rotation, 94-95
 power dissipated in, 64
 precision, 93
 RMA color code, 101-102
 specifications and standards for, 116-117
 temperature characteristics of, 100
 variable-carbon, uses for, 114
 variable wire-wound, 93-95
 water cooling of, 104-105
 wear characteristics of, 115
 wire-wound, 88-92
 Resonance, 72-73
 parallel, 207-213
 series, 203-207
 Resonance bridge, 303
 Resonance-rise method for measuring resistance, 291
 Resonator *Q* values, 719-720
 Resonators, cavity, 718-730
 coaxial-line, 723-725
 cylindrical, 721-722
 rectangular, 720-721
 shunt impedance, 720
 spherical, 722-723
 and wave guides, 684-700

- Retrace ratio, 997
 Reverberation, 763
 R-f amplifiers (*see* Amplifiers, r-f)
 R-f circuits, band-pass, 223-224
 R-f lines (*see* Lines, r-f)
 Ringing, selective, 874-875
 Ripple in power supply, 334
 Rochelle salts, 420
 Room acoustics, 731, 762-768
 Runway localizer, 1124-1127
- S**
- Saw-tooth waves, television, 1000
 Scanner, continuous feed, 1057
 drum, 1056
 Scanning, facsimile, 1052-1056
 television (*see* Television scanning)
 Scanning radiation, conical-, 667
 Scanning yoke, 1036
 Schering bridge, 299
 Scott audio oscillator, 432
 Scratch filters, 201, 207
 Screens, aluminum-backed, 600-601
 flat reflecting, 671-673
 Selectivity, adjacent-channel, 397
 distant-channel, 397
 variable, 855-856
 Selenium rectifiers, 107, 321-322
 Sensation-stimulus relation, 10
 Sheet, electrical, 148
 Sheet radiators, current in, 661-662
 Shielding, and bonding, 884
 and filtering, 827-828
 Shielding tube, 578-579
 Ship antennas, 633
 Shoran, 1121-1122
 Shunt, Ayrton-Mather universal, 259
 Signal coverage, broadcast-station, 986-987
 Signal shaping, 1061
 Silicon crystal mixer, 862
 Silmaul, 150
 Skin effect, 86-87, 118
 on instruments, 262
 Skip distance, 528
 Slide-wire bridge, 292-293
 Solenoid, multilayer, 125
 single-layer, 124-125
 Sommerfeld formula, 526
 Sonne, 1102-1103
 Sound, measuring units of, 8-9
 physiological characteristics of, 8-9
 Space charge effect, 584
 Speech characteristics, 731-733, 851
 Speech-input amplifiers, 921
 Speed control, Hammond brake, 1057
 Sperry air-traffic control system, 1144-1146
 Square-law detectors, 506
 Squelch and noise-limiting circuits, 821-823
 Stabilization, f-m carrier, 957
 Standard-signal generators, 285-286
 Static, 543
 precipitation, 885, 1154-1156
 Static dissipators, wick, 1154
 Static fields, conservative properties of, 32-33
 Steady-state solution, 68-69
 Strip, electrical, 148
 Stub lines, r-f, 245
 Studio technique, 914-922
 Superheterodynes, 776, 799-809
 spurious responses in, 498-499
 Superposition principle, 58, 65
 Superregeneration, 509-512, 863-870
 Susceptance, 71-72
 Susceptance-variation method for measuring
 resistance, 290-291
 Suspensions, coating of, 552
 Switching, lobe, 665-666
 Synchronizing, facsimile, 1062
 television (*see* Television synchronization)
- T**
- T networks, 306-307
 Tank circuits, tapped, 215-216
 Tapers, 93, 114-115
 Tees, wave-guide, 711
 Telemetering and remote control, 828-831
 Telephone, 278
 Teleran system, 1148-1150
 Television, 995-1050
 color, 1039-1048
 color balance in, 1043
 color matching in, 1040-1043
 color systems for, sequential, 1039, 1043-1045
 simultaneous, 1039
 electron-beam deflection in, 1033
 frame-repetition rate for, 999-1000
 horizontal shaping unit in, 1011-1012
 i-f amplification in, 1025-1026
 infrablack region in, 1002
 resolution of, horizontal, 998
 ratio of, 998
 vertical, 997-998
 shading-correction signal generator for, 1007,
 1013
 timing unit in, 1011
 vertical shaping unit in, 1012-1013
 wide-band amplification in, 1024-1025
 Television cameras, 1006
 Television images, contrast of, 1036-1039
 gradation of, 1036-1039
 Television picture tubes, 1030
 contrast in, 1031
 Television power supply, 1033
 Television scanning, 995-1002, 1034-1036
 interlaced, 996-997
 linear, 995-996
 rate of, 1000
 wave forms in, 1000-1002
 Television synchronization, 1034-1036
 pulses of, 1004-1005
 signal separation in, 1028-1030
 Television video amplification (*see* Video amplifi-
 cation)
 Television viewing-distance relationship, 999
 Temperature characteristics of resistors, 100
 Temperature coefficient of resistance, 84-85
 Temperature control of piezoelectric crystals,
 424-425
 Tetrode design, 575-576
 Thermalloys, 149
 Thermionic emission, 545-555
 Thermionic testing, 554-555
 Thermistors, 88, 108-110
 uranium-oxide, 110
 Thermocouple meter, 263-266
 Thermoelements, vacuum, 266
 Thermophones, 279
 Thévenin's theorem, 66
 Thomas's bridge, 299
 Threshold, of hearing, 9
 of pain, 9
 Threshold limiters, 521

- Thyrite, 108
 Time constants, 193
 Time effects, transit-, 584-586
 Timing axis for CR tubes, 284
 Tone control, 201, 207, 823, 1181
 Toroidal iron-core coils, 129
 Tourmaline, 420
 Tower radiators, bonding of, 626-627
 Tracking, 122, 131, 213-215, 798, 808-809
 Tracking error, 834
 Traffic control procedures, CAA, 1136-1138
 Transconductance, 572-573
 Transformer-coupled amplifier, 219, 354-360, 394
 Transformer-coupled r-f circuits, 219-220, 394-396
 Transformers, bridge, 295-296
 color code, 330, 892
 impedance-matching, 367-369, 716
 input, design of, 371
 iron-core power, 161
 low-power, design of, 322-332
 wave-guide, 709-710
 and windings, impregnation of, 343-344
 Transient currents, 188-198
 LCR, 191-193
 LR, 190
 RC, 189
 Transient solution, 69-70
 Transition frequency, 834
 Transitron oscillator, 434-435
 Transmission, distortionless, 25-26
 vestigial sideband, 1023-1024
 v-h-f, shadow loss for, 536
 Transmission-line calculations, 974-976
 Transmission-line speaker, 761-762
 Transmission lines, and networks, 226-253
 r-f, loss in, 985-986
 table of, 240
 and wave guides, 684-686
 Transmitters, arc, 1167
 automatic, 1180-1181
 marine, 1169
 radiobroadcasting, 945
 Transponder systems, 1139-1142
 Traveling-wave tubes, 596-599, 863
 Trigonometric functions, 648-649
 Triode, design of, 567-575
 lighthouse, 860
 negative-anode, 574-575
 Tubes, 545-601
 with beam modulation, 587-596
 beam-power, 579-580
 cathode-ray (*see* Cathode-ray tubes)
 converter, 580-582
 cyclophon, 871
 disk-seal, 450
 of force, 30
 klystron, 587-592
 mixer, 498
 mu-slop, 572
 pentode, 576-579
 picture, 1030
 reactance, 825
 rectigon, 318
 remote cutoff, 579
 shielding, 578-579
 traveling-wave, 596-599, 863
 water-cooled, 966
 Tuners, butterfly, 861-862
 Tungar tubes, 318
 Tungsten filaments, 546-547
 Tuning, inductive, 797-798
 multiple, 615
 permeability, 399, 808
 push-button, 809-812
 Tuning elements, concentric lines as, 402
 Tuning-fork oscillators, 455
 Tuning-fork standards, 255
 Tuning indicators, 820-821, 846
 Tuning stability in r-f amplifiers, 399
 Turner constant-impedance inductance measurement, 160-161
 Turnover point, 926
 Turnstile elements, 663
- U
- Units, and dimensions, 78-82
 rationalized and unrationalized systems of, factors for, 32
 Uranium-oxide thermistors, 110
- V
- Vacuum thermoelements, 266
 Vacuum-tube meters (*see* Electron-tube meters)
 Vacuum-tube oscillators, (*see* Oscillators)
 Variable-area recording, 922
 Variable-density recording, 922
 Varistors, 88, 107
 Vectolite, 150
 Velocity, group, 688
 phase, 688
 Velocity microphone, 902-903
 Velocity modulation, 461
 Vented enclosure, 758, 760
 Vibrating-reed meter, 277
 Vibration galvanometer, 260
 Video amplification, 1015-1021
 1-f compensation in, 1017-1018
 maximum frequency in, 1000
 noise limitations to, 1020
 series-peaking compensation in, 1017-1018
 transient response in, 1019-1020
 Video detection, 1026-1027
 Video modulation, 1021-1023
 Video-signal generator, 1005-1006
 Video wave form, 1002-1015
 Viscalloy, 150
 Vision, characteristics of, 13
 Voltage, 34, 50, 255
 breakdown, of typical dielectrics, 61
 current and energy relations, 68
 and currents, out-of-phase, circuits for, 225
 Voltage-doubling circuits, 317-318
 Voltage-measuring instruments, 272-274
 Voltage ratings of r-f lines, 245-246
 Voltmeters, electrostatic, 273
 Volume compression, 838-840
 Volume control, 918-919
 automatic, 813-820
 Volume-control-resistance curve, 115
 Volume expansion, 838-840
 Volume indicators, 919-921
 Volume range of broadcasting, 900
 Volume resistivity, 84
 Vu, 8, 16, 900-901
 Vu meter, attenuators for, 271
- W
- Wagner ground, 310
 Wave analysis, 284-285

- Wave filters, 246-253
Wave forms, scanning, 1000-1002
 video-signal, 1002-1015
Wave-guide bridge structures, 711-713
 rat-race, 713
Wave-guide connectors, 707
Wave-guide modes, 686-687
Wave-guide tees and junctions, 711
Wave guides, attenuation in, 693-694
 and cavity resonators, 684-730
 characteristic impedance of, 692
 circular, 690-691, 694-697
 conductor losses in, 694
 dielectric, 697-699
 elliptical, 691-692
 fixed attenuators for, 714-715
 mode transformers for, 709-710
 obstacles and discontinuities of, 697-707
 power-carrying capacity of, 692-693
 rectangular, data table of, 689
 and transmission lines, comparison of, 684-686
 waves in, launching of, 707-709
- Wavelength, fundamental, 610
Wavemeters, cavity, 729-730
Waves, electromagnetic, 75-78
Wax preparation, 930-931
Weber-Fechner relation, 11, 13
Weston cells, 255
Wheatstone bridge, 291
Whip antennas, 619-620
Wien bridge, 304
Wien-bridge oscillator, 430
Wire insulation, 139-140
Wireless phonograph attachments, 836
Wiring, color codes for, 892-893
- Y
- Yrneh, 63
- Z
- Zirconium arc light, 1058

