1962 IRE International Convention Record

PART 7

Sessions Sponsored by

IRE Professional Groups on

Audio

Broadcasting

Broadcast and Television Receivers

at

the IRE International Convention, New York, N.Y.

March 26-29, 1962

The Institute of Radio Engineers

1962 IRE INTERNATIONAL CONVENTION RECORD

An annual publication devoted to papers presented at the IRE International Convention held in March of each year in New York City. Formerly published under the titles CONVENTION RECORD OF THE I.R.E. (1953 & 1954), IRE CONVENTION RECORD (1955 & 1956), and IRE NATIONAL CONVENTION RECORD (1957, 1958, & 1959).

Additional copies of the 1962 IRE INTERNATIONAL CONVENTION RECORD may be purchased from the Institute of Radio Engineers, 1 East 79 Street, New York 21, N.Y., at the prices listed below.

Part	Sessions	Subject and Sponsoring IRE Professional Group	Prices for Members of Sponsoring Professional Group (PG), IRE Members (M), Libraries and Sub. Agencies (L), and Nonmembers (NM)			
			PG	M	L	NM
1	8, 16, 23	Antennas & Propagation	$ .70	$ 1.05	$ 2.80	$ 3.50
2	10, 18, 26, 41, 48	Automatic Control Circuit Theory	1.00	1.50	4.00	5.00
3	1, 9, 17, 25, 28, 33	Electron Devices Microwave Theory & Techniques	1.00	1.50	4.00	5.00
4	4, 12, 20, 34, 49	Electronic Computers Information Theory	1.00	1.50	4.00	5.00
5	5, 13, 15, 22, 29, 47, 54	Aerospace & Navigational Electronics Military Electronics Radio Frequency Interference Space Electronics & Telemetry	1.20	1.80	4.80	6.00
6	3, 11, 31, 35, 42, 45, 50, 52	Component Parts Industrial Electronics Product Engineering & Production Reliability & Quality Control Ultrasonics Engineering	1.40	2.10	5.60	7.00
7	30, 37, 43, 51	Audio Broadcasting Broadcast & Television Receivers	.80	1.20	3.20	4.00
8	7, 24, 38, 46, 53	Communications Systems Vehicular Communications	1.00	1.50	4.00	5.00
9	2, 19, 27, 32, 39, 40, 44	Bio-Medical Electronics Human Factors in Electronics Instrumentation Nuclear Science	1.20	1.80	4.80	6.00
10	6, 14, 21, 36	Education Engineering Management Engineering Writing & Speech	.80	1.20	3.20	4.00
		Complete Set (10 Parts)	$10.10	$15.15	$40.40	$50.50

Responsibility for the contents of papers published in the IRE INTERNATIONAL CONVENTION RECORD rests solely upon the authors and not upon the IRE or its members.

1962 IRE INTERNATIONAL CONVENTION RECORD

PART 7 - AUDIO; BROADCASTING; BROADCAST AND TELEVISION

RECEIVERS

TABLE OF CONTENTS

A STEREOPHONIC CERAMIC PICKUP CARTRIDGE FOR 2 GRAM TRACKING

A. L. Di Mattia, E. Kaulins, B. B. Bauer

CBS Laboratories
High Ridge Road
Stamford, Connecticut

Summary

The pickup described will track with a vertical force of 2 grams, on simple but properly designed tone arms. It has a mechanical compliance 6 cm x 10^{-6}/ dyne and an effective mass of about 1.4 mg. Output is 0.35 volt. A combination of optimum rubber properties and viscous damping provide a response free of resonances.

The design is well suited for quantity production for the home phonograph market. Another paper describes successful tracking of this cartridge in commercially available record changers.

In a pickup destined for quantity production, several design factors must be met in order to maintain uniform high performance. It is important that the two ceramic elements be correctly positioned during assembly. In the present design this is done by means of octagonal rubber element mounts which are placed into two semi-octagonal channels on mating interfaces of two plastic housing parts.

In any two-element stereophonic pickup design, a "coupler" is needed to join the two elements to a single stylus arm. Design of the coupler has important bearing on distortion, output level, channel separation, and mechanical impedance. Coupler requirements in the present design have been met by proper configuration and choice of material.

Another important design factor is the means provided for pivoting the stylus lever arm. For lateral motion, at the stylus, compliance must be high for lateral motion in two degrees of freedom. Rotational compliance, about the leveraxis, must be low in order to avoid undesirable rotation of the lever arm. There must be no lateral freedom at the pivot point. Longitudinal compliance of the lever arm must be low in the direction of drag from the record groove to avoid frequency modulation effects.

The stylus lever mounting arrangement described achieves these design requirements in a simple assembly. A light tube, comprising the lever arm, is flattened at its pivoted end in a manner to produce tapered shoulders. These shoulders are brought to bear against a hole in a small rectangular metal enclosure which is filled with a plastic material having viscoelastic properties. The potting arrangement damps any lever resonances and provides a small restoring force to position the lever for engaging the coupler during change of styli.

Introduction

Design requirements for phonograph pickups are well known. To restate them briefly, a pickup must convert mechanical energy imparted to a stylus from the record groove into corresponding electrical energy. Mechanical impedance at the stylus point must be as low as possible in order to minimize the perpendicular tracking force needed to play the record and to reduce the consequent wear of stylus and groove. The electrical output must be as free as possible of any form of distortion of the groove information. To a stereophonic system one must add another important requirement: the pickup must respond independently to signals of the two stereophonic channels. Ideally, each pair of pickup output terminals should contain signals derived from its associated groove channel only, and not from the other. The ratio of magnitudes of signals derived from successive equal modulation of the wanted and unwanted channels is called "channel separation," and is commonly expressed in decibels.

It is historically interesting and significant that the tracking force started at several ounces, or over 100 grams, in the early days of phonograph reproduction and has now dropped to about 6 to 9 grams in commercially built home phonographs. The more costly pickups currently used by high fidelity enthusiasts track at about 2 grams.

This paper describes a high fidelity stereophonic ceramic pickup cartridge which is capable of 2 grams operation on commercial record changers of good design and is intended primarily for the home phonograph field. In addition to meeting exacting performance standards, this cartridge produces sufficient output voltage for use with home phonograph amplifiers. An accompanying paper describes a tone arm of unique design, which protects records and overcomes problems attending tripping of record changers at 2 grams tracking force.

Design Principles

Many transducer systems are available to the pickup designer, and much has been written about their relative merits. Pickup design is essentially a mechanical problem. It may be said that the best transducer system is the one whose design leads to the closest fulfillment of the various important performance requirements. The piezo-electric transducer has long emerged as the best system for pickups intended for home phonographs because of its high output and convenient electrical equalization. The pickup to

be described employs piezo-electric ceramic elements of the lead zirconium titanate type.

A ceramic element has a high mechanical impedance which must be converted to the required low impedance at the stylus point. Ordinarily this is done by mounting the element in elastic supports, by compliant connection to the stylus, or by a mechanical lever between stylus and ceramic element. Any of these may be used singly or in combination. The lever is to be preferred because it is a mechanical transformer; that is, it lowers mechanical impedance as the square of the lever ratio, while electrical output decreases as the first power. However, high order lever ratios are difficult to design and control in manufacture. Practical limits are ratios of about 3 or 4 to 1.

Most ceramic stereophonic pickups use two transducer elements, one for each of the two program channels. Since there is but one stylus, some form of hinge or coupler is needed to connect each ceramic element to the stylus for independent motion. Ideally, if a coupler transmits each channel information to its respective transucer element only, separation will be infinite. In some stereophonic pickups, separation decreases with increasing frequency and sometimes is reduced to zero or even takes on negative values. At low frequencies, channel separation is determined by the geometry of the mechanism. At high frequencies, slight unbalances caused by lack of symmetry become important. In the pickup to be described, excellent separation is maintained throughout the audio frequency spectrum.

Design of the Pickup

Mechanical impedance transformation from ceramic elements to stylus is accomplished by a combination of compliant element mounts, a compliant coupler and a lever between stylus and coupler. Emphasis has been placed on a design that would permit precise placement of elements using modern line assembly techniques. Ceramic element mounts are octagonal in cross-section, as shown in figure 1-A. The octagon configuration results in a unique combination of angles. Ceramic elements are placed parallel to a pair of sides, perpendicular to another pair, and 45° to the remaining 4 sides. Both mounts are clamped between two plastic shells containing complementary semi-octagonal grooves with small interference fit between grooves and mounts. This provides the required 45° orientation. Another advantage of the octagon shape is that two sides of the mount are perpendicular to the shell interfaces and therefore are aligned with the direction of assembly. Once the parts are correctly assembled by means of an assembly fixture, there is little likelihood that any distortion of element position will occur. This arrangement is depicted in perspective by figure 1-B. The material for the mounts possesses some visco-elastic properties which provide partial damping of the system.

During development much attention was directed to the coupler. Several materials and some novel configurations were evaluated. The final design is the configuration shown in figure 2. Material is a visco-elastic formulation similar to the one used for the element mounts. Two slightly undersized cavities are provided in the coupler for engaging the free ends of the ceramic elements. Linkages from each element converge to form a yoke which will engage the stylus lever arm. The coupler is fairly compliant and contributes to the final compliance of the pickup. The stylus lever arm is placed at the intersection of the two perpendiculars to the elements, as shown. Transverse deflections of the linkages cause torsional forces at the elements, which do not generate any output. Compliance of the linkages is relatively high in the transverse direction, thus further reducing any undesired "cross-talk" forces at the non-driven element.

The requirements for a stylus lever arm in a stereophonic pickup are described with the aid of the diagram in figure 3. The lever is pivoted at point A, carries a stylus which engages the groove at point B, and engages the coupler at point C. The force transformation ratio is (l_2/l_1), and the impedance transformation ratio is $(l_2/l_1)^2$.

The lever must perform the following functions:

(1) Any motions of the stylus in the surface (b) must be translated into equivalent motion at the surface (c).

(2) The stylus extends a distance from the lever arm to provide clearance between the lever and the record. This creates a moment arm that tends to produce rotational motion around the axis A-B, instead of the desired translational motion at C. For this reason, the lever itself must be axially stiff and the pivot at A must prevent rotational motions.

(3) The frictional drag between the stylus and the groove creates a force pulling at the pivot. This drag varies with groove modulation so that there is tendency to produce an effect known as longitudinal distortion if the lever is permitted to move. Therefore, the pivot at A must prevent motions in the longitudinal direction.

Various attempts to solve these multiple problems led to the design depicted by figure 4. Two stylus levers are provided back-to-back in a turret arrangement which is rotated 180° to change styli. A turret pivot post is attached to a formed metal part comprising a small box and two extensions that surround the lever arms to afford some protection. Both ends of the box are pierced with "key-hole" openings as shown. Lever arms are made of thin walled aluminum alloy tubing, one end of which is flattened and pierced to receive a stylus. The other end is also flattened in the form of a spade with flaring shoulders between tube and spade.

In manufacture, the turret is placed in a production fixture. The lever arms are inserted in their respective openings by aligning the spades and key-holes as shown. Both tubes are then rotated 90° to make them captive. In the production fixture, means are provided for holding the lever arms in place. A small spring force (in the fixture) pulls the lever arms outwardly so that

the flaring shoulders are in contact with the inner walls of the box. As the next operation, the box is filled completely with a vinyl dispersion in a liquid plasticizer. At this stage, the material is a fluid of medium viscosity. After a short period at an elevated temperature, the compound fuses to a soft but durable elastomer. A small hole is provided through the spade end. This becomes filled with the elastomeric compound and provides a large increase in rotational stiffness about the lever axis without decreasing the transverse compliance at the stylus. The entire double stylus assembly is now locked together in a position determined by the production fixture.

Figure 5 depicts the cartridge in vertical cross-section. The octagonal channels extend beyond the rubber element mounts and terminate in a larger opening that surrounds the coupler. The pickup is only partially damped as it has been described to this point. Complete damping is obtained by injecting a measured amount of stable non-creeping grease around the elements and within the octagonal channels.

Figure 6 is a photograph of the complete pickup with a detached turret assembly shown below the cartridge. Replacement of the stylus assembly is a simple fingertip operation without tools.

Figure 7 depicts frequency response and separation for both left and right channels. These characteristics were automatically plotted using the CBS Laboratories STR-100 stereophonic test record in conjunction with a General Radio type 1521-A graphic level recorder. From low frequencies to 1 kc, pickup response is flat after correcting for the load resistance used in the measurements. Above 1 kc, response generally follows the constant velocity characteristic of the record with a slow and gradual rise to about 15 kc after which there is a slight roll-off to 20 kc. Some values of separation taken from the curves are: 25 db at 100 cps, 35 db at 1 kc, 19 db at 10 kc, 18 db at 15 kc, and 14 db at 20 kc. Compliance is 6×10^{-6} cm/dyne, effective mass at the stylus is 1.5 milligrams. Output is 0.35 volt at 1 kc for an RMS velocity of 3.54 cm/second.

Acknowledgements

These cartridges are currently being manufactured by the Zenith Radio Corporation where they will be used with the Zenith "Micro Touch" tone arm. The authors are indebted to Mr. Donald Knight, Chief Audio Engineer, and other members of the Zenith staff for their participation in the production engineering stages of this project. The photographs and performance data given herein, are based on production cartridges supplied by Zenith.

The authors also acknowledge early design contributions by Daniel P. Doncaster and George W. Sioles of CBS Laboratories, and Johann Van Leer formerly of the Laboratories.

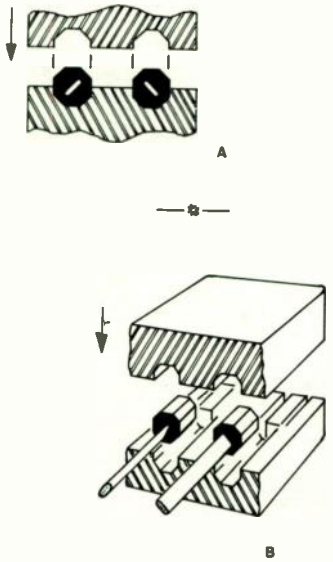


Fig. 1.

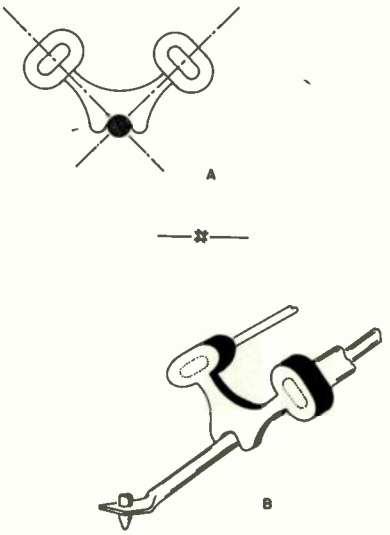


Fig. 2.

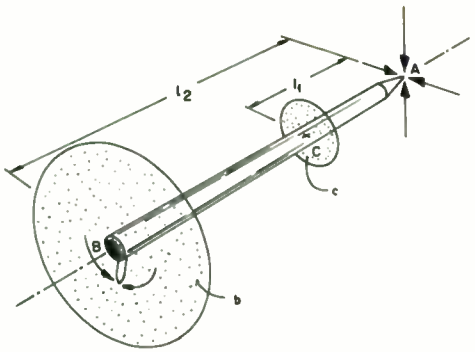


Fig. 3.

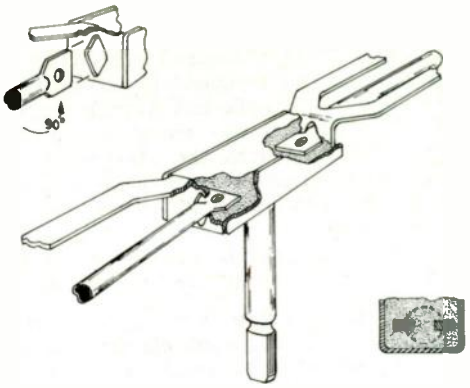


Fig. 4.

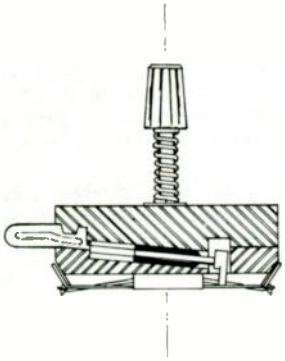


Fig. 5.

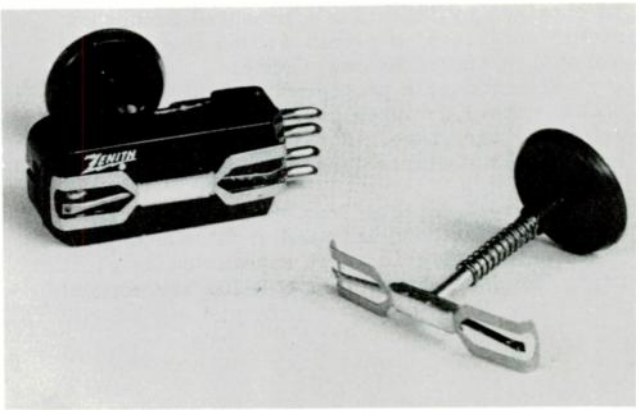


Fig. 6.

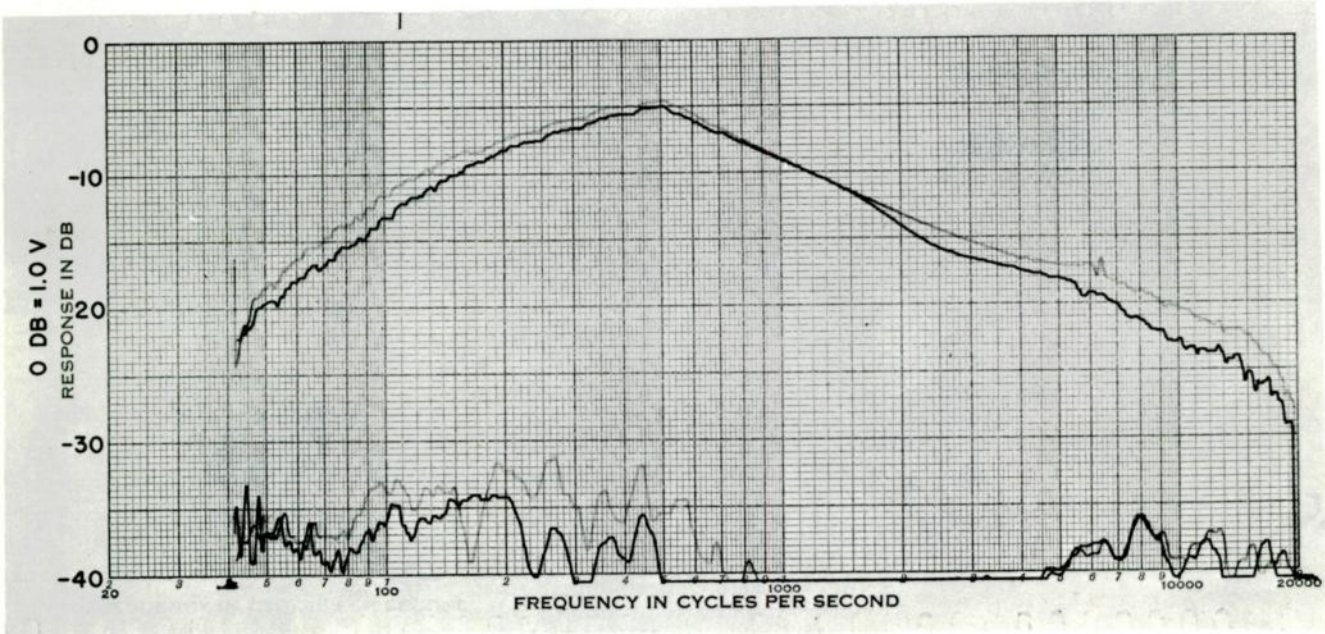


Fig. 7. Output-0.35 Volts/3.54 Rms Velocity, a 1 kc.
Compliance-6×10^{-6} Cm/Dyne
Effective Mass-1.5 Milligrams
CBS Laboratories Test Record Str-100
Shunt Capacitance-125 MMF
Resistive Load-1 Megohm
Tracking Force-2 Grams

A STEREO TONE ARM FOR TRACKING AT 2 GRAMS ON A RECORD CHANGER

G. W. Sioles
B. B. Bauer
CBS Laboratories, Stamford, Connecticut

INTRODUCTION

It has long been recognized that very low tracking forces in phonograph pickups will minimize wear of the record and styli. Another advantage of low tracking force is the reduction in wavelength loss. Two factors tend to place a limit on the minimum tracking force which can be used. One is the mechanical impedance of the pickup and the other is the stability of the pickup-tone arm system when subjected to jarring. Tracking force of approximately 2 grams is usual in high-priced manual players which are handled carefully by relatively skilled people. However, in most home phonographs tracking force is found to be above 6 gms. to insure stability when subjected to vibration. This paper describes a tone arm configuration which was developed to allow tracking at 2 grams in automatic record changers with a low mechanical impedance pickup described in a companion paper. As an additional benefit the new tone arm protects records by preventing damage due to accidental scratching or dropping of the tone arm.

Description of Tone Arm

Figure 1(a) shows a schematic representation of the tone arm design. The pickup is held in a carriage which is pivoted about a horizontal axis near the stylus tip. A trailing spring is attached to the overhung portion of this carriage and provides the force required to take up the net unbalanced force at the stylus tip, as well as isolate the cartridge from the tone arm. The vertical pivot for the arm is located behind the horizontal pivot and a weight is placed at the rearmost part of the arm to statically balance it in the horizontal plane. A second spring, disposed as shown in the sketch, is required to reduce the unbalanced vertical force at the stylus tip to 2 grams.

Equivalent Circuit Representation of the Mechanical System

Figure 1(b) shows the analogous circuit of the mechanical constants of the tone arm for the vertical plane. The first inductance represents the mass of the cartridge and its carrier, and the second inductance depicts the mass of the tone arm; both of these masses being referred to the stylus tip (all rotational mechanical reactances are referred to the stylus tip as either compliances or masses). Three capacitances are shown; the first is analogous to the compliance of the stylus, the second represents the compliance of the first spring, and the third, the compliance of the second spring. The mechanical impedance plot for the circuit is shown in figure 1(c) on the assumption that the mechanical resistances are small. It may be noted that there are two resonances and two anti-resonances. It is interesting to compare this impedance characteristic with that of the conventional arm with respect to stability considerations. In a conventional arm there would be only one anti-resonance frequency defined by the stylus compliance and the arm mass occurring at an intermediate frequency to the ones shown. In the arm under consideration, the anti-resonance occurring at the higher frequency is caused by the compliance of the stylus and the mass of the cartridge and its carrier, the latter modified by the other elements in the circuit. Since the compliance of the first spring in the equivalent circuit is much higher than the stylus compliance, and the mass of the arm is much larger than the mass of the cartridge and its carrier, the remaining part of the circuit is effectively uncoupled. Thus, for all practical purposes, the upper impedance maximum is a result of the stylus compliance with the mass of the cartridge and carrier. The impedance at the anti-resonance frequency is equal to the mass divided by the product of the compliance and resistance, for small values of resistance. Because of the low mass of the cartridge plus carrier the impedance at this frequency is considerably lower than it would be otherwise. Velocities arising from jarring of the equipment produce smaller forces at the stylus tip than would otherwise be the case with an attendant increase in tracking stability. Naturally, the frequency response would be affected if this anti-resonance were allowed to occur in the desired transmission range and a high compliance stylus is needed to produce a satisfactorily low anti-resonance with the small allowed mass.

We now consider the lower maximum in the impedance which arises from the parallel resonance of the trailing spring compliance with the tone arm mass, whose reactance is reduced slightly by the compliance of the second spring. The compliance of the second spring is greater than that of the first, so that the anti-resonance we are observing is effectively that between the mass of the tone arm and the compliance of the

trailing spring. This anti-resonance is considerably lower in frequency than that defined by the stylus compliance and the arm mass because, as previously noted, the first spring has a higher compliance than the stylus compliance. The impedance at this anti-resonance is inversely proportional to the compliance and with unchanged arm mass, the impedance at anti-resonance is lower than would normally be the case in accordance with the relation noted earlier. In short by replacing the single impedance maximum of the conventional tone arm by two maxima with lower mechanical impedance, the force tending to throw the stylus out of the groove when any part of the system is subjected to impulsive velocities has been reduced. It will be noted that whereas the impedance tends towards zero at the higher frequencies it approaches infinity as the frequency goes to zero. This might raise the question as to whether the system is truly stable when subjected to shocks. The answer to this is that the reactance depicted is extremely small and does not become large until a very low frequency - in fact so low that the expected spectrum of disturbing amplitudes lies above the range where forces would be developed sufficient to cause mistracking. This has been verified by experience.

Alignment of Pivots

With the cartridge pivot so close to the stylus, the playing of a warped record will result in frequency modulation or "wow", which by proper design is made negligible. The problem is not severe for a gentle warp since there is relatively little differential motion between the tone arm and the cartridge. This can be seen by viewing the equivalent circuit for the system as comprising the compliance of the trailing spring C_1, in parallel with the mass of the tone arm, m_A. For a gently warped record this circuit is excited below the resonance frequency and the tone arm motion is much greater than that of the cartridge. If the warp is more abrupt, approaching an impulse, the equivalent circuit is excited at its resonance frequency and both the cartridge and tone arm will be displaced in proportion to the relative resistance in each branch. The location of the cartridge pivot in relation to the stylus, and the angle the line joining these points makes with the record become important if "wow" is to be prevented. This is clear from the geometry. Since the stylus moves on an arc, the component of motion in the direction of the groove is increased when the pivots are not on a line parallel to the record. In practice, a slight downward angle is used so that the pivot will clear the record.

Scratchless Feature

The method of elastically mounting the cartridge produces a performance advantage in regards to protection of records and fragile stylus tips. In a conventional arm, an accidental downward thrust on the arm produces displacement of the stylus and a resultant force proportional to this displacement which may be sufficient to plastically deform the record material. Frequently such accidents produce a lateral motion of the arm and result in scratching of the record; and in more unfortunate cases the stylus tip is broken. Because the compliance of the elastic mounting is larger than that of the stylus, the force between stylus and record is lessened for such occurrences with a corresponding reduction in the amount of damage done. In the design described the maximum force which can exist between the stylus and the record is approximately five grams, at which point the cartridge has retracted sufficiently into the arm to no longer be vulnerable.

An example of the relative damage to a record with this arm compared with conventional designs can be seen in Figure 2. These photomicrographs show the width of groove embossed into the disc when the arm was deliberately caused to "skate" across the record.

ACKNOWLEDGMENT

The authors express their thanks to I. Delin, A. L. DiMattia and D. Doncaster for their technical contributions in the development of the device.

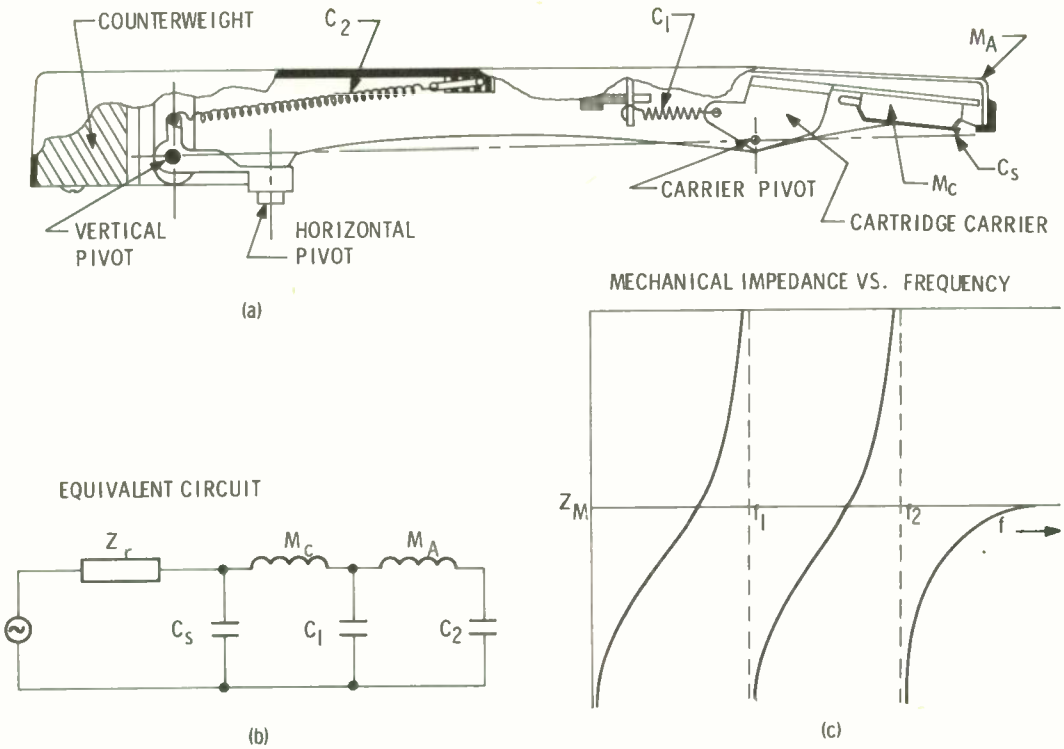


Fig. 1.

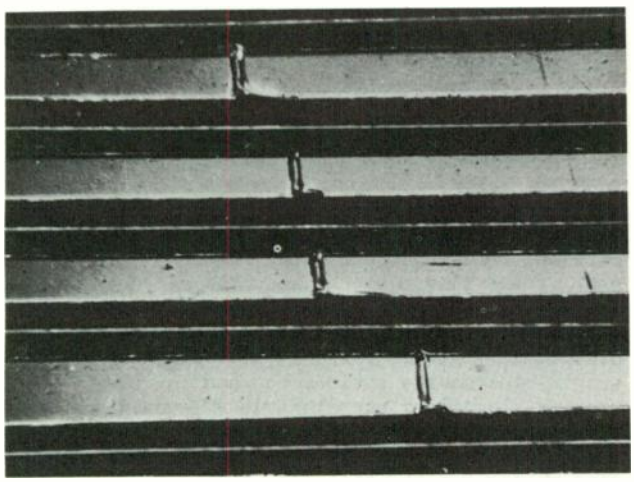


Fig. 2(a).

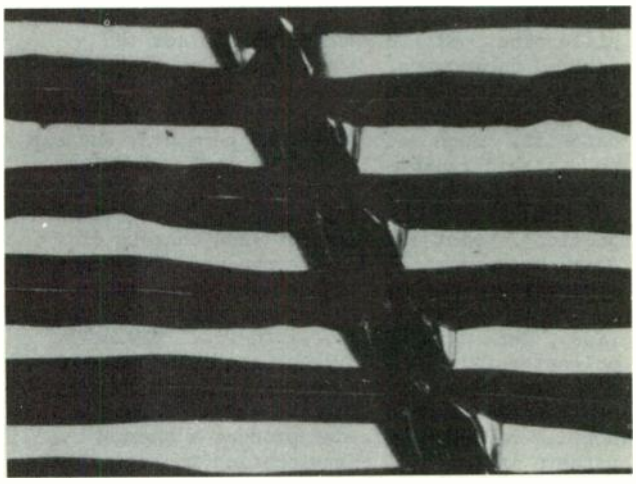


Fig. 2(b).

A SYSTEM OF ELECTROSTATIC RECORDING

D. E. Richardson, J. J. Brophy and H. Seiwatz
Armour Research Foundation
Chicago 16, Illinois

and

J. E. Dickens and R. J. Kerr
Film Department
E. I. Du Pont de Nemours and Company
Wilmington, Delaware

Abstract

A new system of electrostatic recording in which electrical charges corresponding to the signal are injected into a thin insulating tape record has been investigated. The recorded charge pattern is produced by the application of a.c. or d.c. bias potentials to the recording electrodes, in addition to the signal voltage, and subsequently stabilized by treating the tape surface with ions to neutralize excess charge. Electrostatic recordings with extrapolated lifetimes in excess of several hundred years have been obtained.

Introduction

Previous electrostatic recording techniques have generally employed either electrical polarization of the record medium, based on electret[1] or ferroelectric[2] properties, or surface charging of an insulator.[3,4] These methods yield electrostatic records which lack the desired permanence. A new system of electrostatic recording reported here utilizes charge penetration into an insulating recording material. Such recordings are permanent and can be replayed repeatedly.

Early experiments showed that charge injection is most advantageous when charges of equal and opposite sign are produced at points opposite one another through the tape. This is accomplished by drawing certain plastic films (i.e., Teflon FEP fluorocarbon film, "Mylar"* polyester film, and polyethylene film) between a knife edge electrode and a resilient conducting backing electrode. Bias and signal voltages applied to the electrodes produce a charge pattern in the film corresponding to the signal which may be played back inductively by passing the film through a similar electrode structure. Most of the results reported here refer to recording systems using "Mylar" tapes 0.25 inches wide and either 0.25 or 0.50 mils thick.

*Du Pont's registered trademark for its polyester film.

Recording Techniques

With the simple recording and playback arrangement of Fig. 1(a), little charge is injected into the moving tape until the electrode bias potential exceeds a threshold value as shown in Fig. 2. This motional tape current, which is a linear function of tape speed demonstrates that electrical charge corresponding to the electrode potential is delivered to the tape. The signal-to-noise ratio and other properties of such a recording are measurably improved by subjecting the tape to a reverse polarity prebias treatment with a prior knife edge and backing electrode combination. The subsequent motional tape current at the second electrode pair is labeled "-1000 volts prebias" in Fig. 2. In the absence of signal the net result of the prebias-bias treatment is an essentially neutral tape, for properly chosen values of electrode potentials. Signal voltages are applied only to the second electrode combination and result in a charge pattern interior to the tape corresponding to the applied signal. It is possible to erase this charge pattern by subsequently re-recording the tape.

Experiments have shown that maximum signal-to-noise ratio, minimum playback distortion, and maximum record life are obtained when the prebias electrode current is equal to the bias electrode current, which can be realized by proper adjustment of the bias and prebias voltages according to Fig. 2. The proper division of voltages may be obtained automatically through the use of the circuit in Fig. 1(b). Here the voltage of the single bias source divides between the bias and prebias electrodes such that $I_P = I_B$ on the average while the signal voltage is effectively applied only to the bias electrodes. Representative values for the bias potential and signal level in this circuit are 1500 volts d.c. and 100 volts rms, respectively.

Playback of these electrostatic records is simply accomplished by drawing the tape between a knife edge and backing electrode combination as indicated in Fig. 1(a). Tape signal charges induce potentials on the playback electrodes which may be amplified by a conventional electronic amplifier. The magnitude of the signal depends upon the load resistance R_L and is of the

order of 40 millivolts for $R_L = 10^6$ ohms. Typical playback frequency response spectra and wideband noise levels are shown in Fig. 3. In the figure the noise data points refer to wideband noise (10 cps to 20 kc bandwidth) taken immediately before and after each signal response point on adjacent regions of the tape record.

An effect similar to the bias-prebias combination in d.c. bias recording is achieved through the use of a.c. bias at a single pair of opposed knife edges, as sketched in Fig. 1(c). Recordings result which have a larger signal-to-noise ratio than for d.c. bias, as indicated in Fig. 4. Present results also indicate less second harmonic signal distortion in the case of a.c. bias.

Ion Neutralization

The recorded tape is treated with an atmospheric ion bath prior to spooling after recording and after each playback. This ion neutralization treatment improves recording permanence, greatly reduces noise due to random triboelectric charges on the tape, and reduces layer-to-layer print-through during storage. Present evidence indicates that the function of these ions is to neutralize uncompensated charge on the tape. The ion bath is produced by a high voltage corona discharge from pointed electrodes located between the recording electrodes and the take-up spool as indicated in Fig. 1(b). A similar ion source between the feed reel and the electrodes may also be employed, but is often unnecessary. The ion bath neutralizes net charge on the record medium produced by triboelectric charging or the recording process and is applied each time the tape passes through the electrodes.

Noise

Noise in the electrostatic recording system is due to random electrical charge patterns in the tape and to recorded noise produced by the recording process. The playback noise spectrum of virgin tape, Fig. 5, indicates a very large low frequency component, presumably the result of triboelectric charging during handling and spooling. This is greatly reduced by ion neutralization as also shown in Fig. 5. The noise spectra of neutralized tapes are quite reproducible and may be interpreted in terms of a stable random array of charged spots produced by air breakdown at the tape surface because of triboelectric charging. This charge pattern is altered by the application of bias and prebias voltages as indicated by the third spectrum of Fig. 5.

The recorded noise spectrum also can be analyzed in terms of a random array of (much smaller) charged spots produced by the recording process. These spots are in agreement with the observed spark-like character of the gas discharge at the electrodes during recording. The total wideband recorded noise level is greater than that of neutralized tape, although, as the spectra indicate, disagreeable low frequency rumble is much reduced. Therefore, d.c. electrostatic recordings are recording-noise limited with regard to signal-to-noise ratio. Under a.c. bias the character of the recording discharge is such that the spark-like character is reduced and the signal-to-noise ratio is improved as implied in Fig. 4.

Resolution

The linear resolution of electrostatic records is influenced by the effective electrical width of the electrodes on both recording and playback. It is possible to analyze response curves such as in Figs. 3 or 4 in terms of an effective electrode width quite satisfactorily. A minimum wavelength of the order of 0.7 mil is obtained for the simple knife edge playback electrode.

Playback resolution can be improved through the use of a shielded blade electrode construction in which a center pickup electrode is shielded on each side by a close-fitting electrostatic shield. This configuration restricts the region of tape effective in inducing signals on the active electrode and results in a playback electrode having a resolution equivalent to the distance between the shields. In application it is convenient to electrically drive the shields through the use of a cathode-follower input stage in order to reduce the capacitive loading effect of the shields. The improved resolution obtainable is indicated by the curves of Fig. 6 which were taken under identical conditions except for the size of the electrode gap between the shields.

Resolution limitation in recording appears to be associated with the gas discharge at the recording electrodes. The finite size of the charged spots produced in d.c. bias recording and the spread of the discharge away from the knife edge tend to introduce a minimum wavelength. In the latter regard, change in the ambient atmosphere is known to influence resolution to some extent.

Recording Life

One of the outstanding features of this new electrostatic recording system is the long record life, which is attributed to the particular charge distribution produced in the tape by the recording process. Typical signal and noise decay characteristics are shown in Fig. 7 for a tape stored under room ambient conditions. For a period of time following recording the signal and noise decay together, preserving the signal-to-noise ratio, while longer storage increases the noise somewhat, possibly due to a noise print-through effect. The signal decay may be reasonably represented by a power law as shown on the curve with an exponent of 0.3. Values of this exponent as large as 0.7 have been observed

for recording prepared under non-ideal conditions, for example when bias and prebias currents are not equal.

Tapes stored under low relative humidity conditions have even longer lifetimes than indicated in Fig. 7. At low humidity, exponents as small as 0.07 are observed. The life of such recordings is clearly of the order of hundreds of years. The other aspect of record life, multiple playback, appears to be affected by the ion treatment. Recordings given a suitable neutralization treatment can be played back hundreds of times with relatively small diminution of signal.

Summary

Based on these results it appears that the new electrostatic recording system derives its performance characteristics from the specific charge distribution pattern produced in the record medium by the recording process. The effect of ion treatment is to help preserve this pattern through neutralization of net surface charge produced by triboelectricity. The recording noise limitation on signal-to-noise ratio in d.c. recording may be removed by a.c. bias. Linear resolution on the record is a function of both recording and playback processes but the limitation due to the latter may be largely eliminated by the proper electrode configuration. Excellent signal life, both with respect to storage time and multiple playback, are attainable. Further improvements in the system performance will result from a detailed understanding and control of the gas discharge at the recording electrode.

References

1. R. E. Rutherford, U. S. Patent No. 1,891,780, Dec. 20, 1932.

2. C. F. Pulvari, U. S. Patent No. 2,698,928, Jan. 4, 1955.

3. F. Gray, U. S. Patent No. 2,200,741, May 14, 1940.

4. V. C. Anderson, Rev. Sci. Instr. 28, 504 (1957).

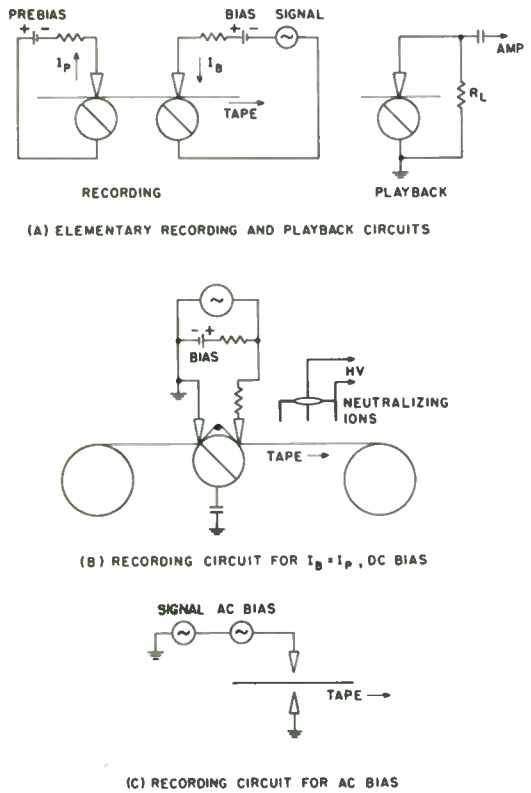


Fig. 1. Electrostatic recording and playback circuits.

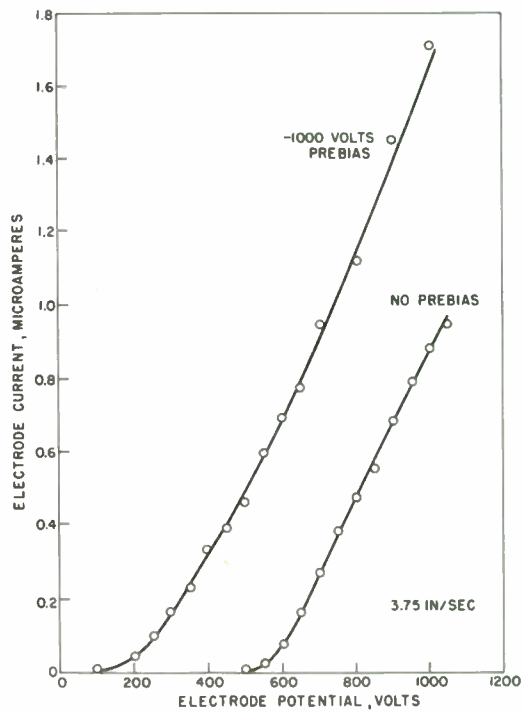


Fig. 2. Recording current characteristics for 1/4 mil thick, 1/4 inch wide "Mylar" at a tape speed of 3.75 inches per second.

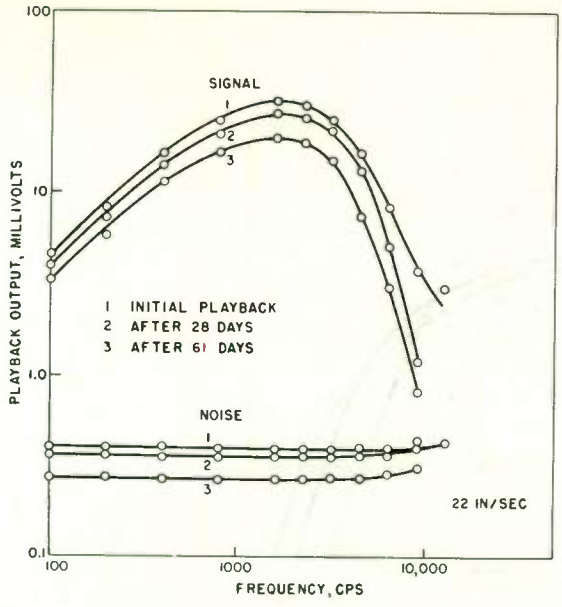


Fig. 3. Frequency response curves for d.c. bias electrostatic recording at 22 ips. Indicated noise levels are total wideband noise with bias but without signal taken immediately before and after signal response readings.

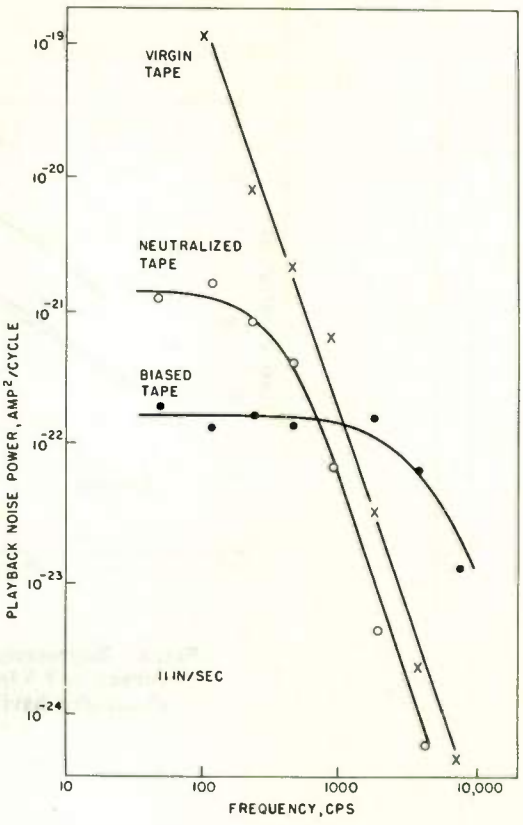


Fig. 5. Playback noise spectra at 11 ips.

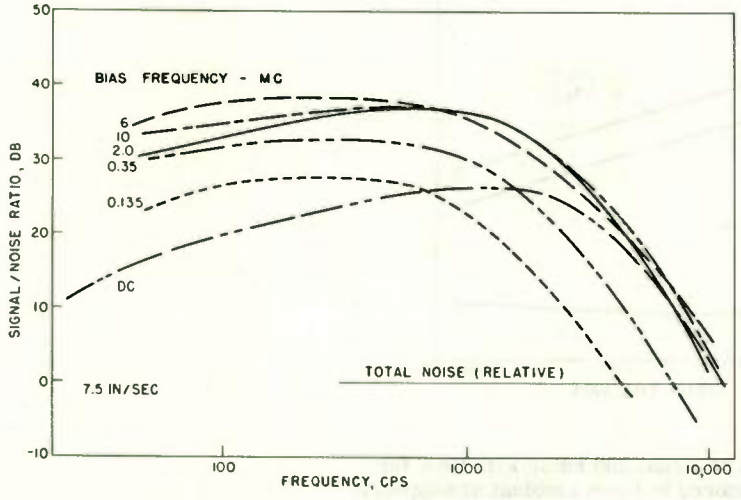


Fig. 4. Comparison of frequency response curves at 7.5 ips for various a.c. bias frequencies and for d.c. bias.

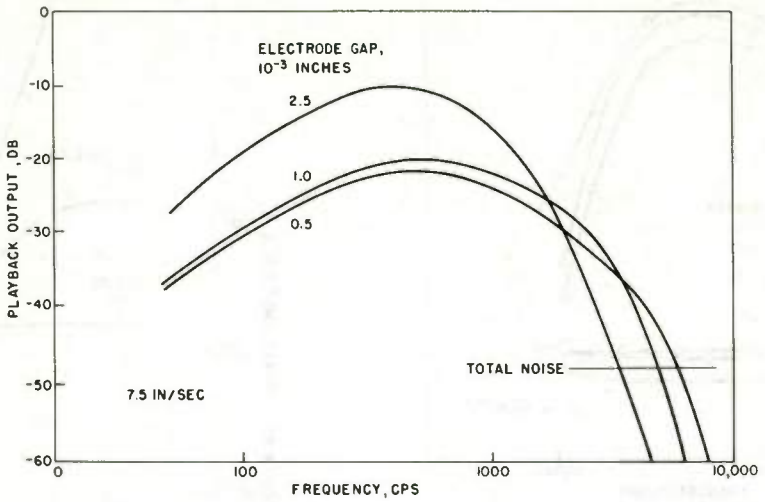


Fig. 6. Representative playback frequency response curves at 7.5 ips obtained with sandwich playback electrodes having various gap widths.

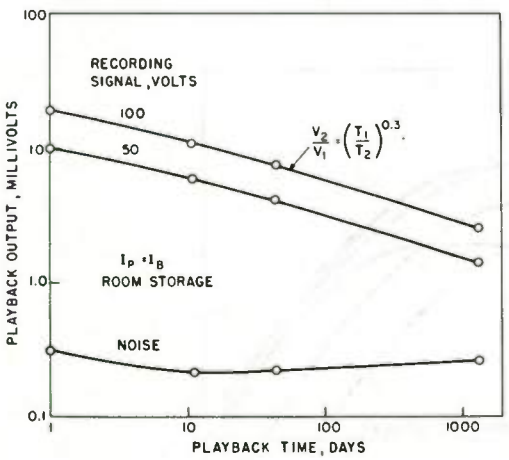


Fig. 7. Variation of signal and noise with time for "Mylar" tape stored in room ambient atmosphere.

A NEW AUTOMATIC LEVEL CONTROL FOR MONOPHONIC AND STEREOPHONIC BROADCASTING

Arthur Kaiser
Benjamin B. Bauer
CBS Laboratories
Stamford, Connecticut

It is a well-known fact that radio listeners tend to favor a station which comes in "loud and clear". This trend has become especially important because of the large number of portable and automotive receivers which operate in noisy surroundings. Since the AVC of radio receivers tends to equalize the carrier strength prior to detection, high modulation level is required to attract the listener. This paper describes a new automatic level control amplifier, called the Audimax, which maximizes the level of modulation in accordance with accepted broadcast studio control practices, but which is not subject to the limitations of human reflexes and span of attention. This is especially important as programming complications have tended to augment the number of tasks performed by studio engineers.

F. M. multiplex stereophonic broadcasting imposes additional problems since two channels of audio are now involved and a specific relationship must be preserved between them. This is further complicated by the fact that modulation levels are determined by the instantaneous sum of these two channels.

General Requirements for Audio Level Control

Using an accepted volume indicator and his hearing, a conscientious technician is required to:

1. Prevent excess levels from overdriving the audio system by a rapid reduction of gain.

2. Offset a reduction in the average audio level by an increase of gain at a modest but steady rate.

3. Recognize that pauses in the audio of several seconds' duration require no alteration of gain.

4. Recognize that occasional excessive audible peaks that are out of context with the mean level will be handled by the peak limiter feeding the transmitter or recording amplifier and require no manual change of gain.

The Prior Art

Since variable gain program-level amplifiers have been available for several decades, the question may be asked, how well do they fulfill the objectives listed above? Let us consider, therefore, a conventional variable gain amplifier which reduces gain above a critical signal input with a speed of a few milliseconds, and upon cessation of the input recovers its gain in a manner which may be adjusted at will. As a typical example, let us assume that gain reduction takes place in 15 milliseconds and 63% of gain recovery in about two seconds. To provide an adequate range of gain control the amplifier is adjusted so that normal input level produces 10 db of gain reduction.

With regard to the first requirement indicated for audio level control, conventional devices provide adequate performance because of their "attack" speed. However, with a recovery speed as indicated, considerable gain increase accompanies the "valleys" associated with speech or singing. This gives rise to the phenomenon known as "pumping" where the level of background sounds or accompaniment rises and falls with variations in level between syllables or groups of syllables. By employing a recovery time of 10 seconds or longer, this effect is alleviated but another problem arises, later to be described.

Referring to the second requirement, conventional devices can provide adequate gain increases to accommodate reduced levels provided the recovery time is sufficiently short. Otherwise a low level caused by switching program sources might produce an extended period of inadequate gain. This requirement, therefore, conflicts with the necessity for eliminating "pumping."

The third requirement must be examined from two aspects: first, brief lapses of audio of about 2 seconds' duration caused by pauses in speech to punctuate sentences or paragraphs or to take a breath and second, pauses of greater duration for dramatic purposes. In the first case, conventional variable-gain amplifiers with short recovery time tend to exaggerate the sound of breath intake immediately preceding the resumption of speech in addition to raising the level of background quite noticeably. In the second case, the full gain of the system would be restored with possible exaggeration of background; for example, motion picture sound tracks. If a long recovery is substituted, the system will then not accommodate a drop in level adequately.

A studio technician will generally have little difficulty in meeting the fourth requirement since his speed of response is inadequate to effect a quick gain reduction. However, conventional variable gain amplifiers operating with long recovery times will cause the program to suffer prolonged gain reduction on occasional short "bursts" of sound. The classic case is that of a pistol shot. Audio immediately following is attenuated for several seconds.

By way of summary, a so-called "conflict of interests" exists with reference to the recovery characteristics of conventional variable gain devices. If too short, we experience "pumping" and the degradation of dynamic fidelity caused by an alteration of the peak-to-valley ratio. If too long, the system will not restore the gain required for a drop in level with adequate speed and it becomes vulnerable to having "holes" in the audio caused by high level short bursts.

This situation has been partially resolved by the inclusion of so-called "dual time - constants." In effect, this is a compromise solution. The use of expander modes to prevent an increase of background level during prolonged pauses is not an altogether satisfactory solution to the problem since it introduces a severe discontinuity in the input-output function, and where the signal level is just beyond the "verge of limiting" the severe gain reduction is as apparent as a gain increase would be without it.

Finally, the listener experiences a feeling of instability caused by the tendency of the gain to follow the contour of the signal continuously in contrast to the solidarity experienced with manual control.

The Audimax System

Figure 1 shows a block diagram of the Audimax system. The main signal channel consists of a high-quality program level amplifier with V_3 as a variable-gain stage to which the control voltage is applied and V_4 as the output stage. The output of detector 1 is stored in a memory from which the comparator extracts information regarding the program content for the preceding 10 seconds. A decision is then made by the comparator based upon the memory content and the output of detector 2 as to whether a gain change is required. If so, this will be effected within 1 to 2 seconds if the gain is to be increased and in less than 15 milliseconds if decreased. Detectors 1 and 2 are differently "weighted." Each responds to those properties of the program signal with which it is concerned for the purpose of enabling the comparator to make a proper decision.

Audimax functions in a four-dimensional domain whose coordinates are input level, output level, memory and time. Figure 2 is a two-dimensional representation of the system gain versus input level. This diagram illustrates the operation of the memory device. The gain at any time is dependent on both the weighted input level and the program content during the preceding 10 seconds. For example, for a range of input levels from A to B, the gain of Audimax remains on a stable "platform" C-D as determined by the comparator. If the range of input levels shifts to a new region, the gain "platform" is quickly re-adjusted to the new value required to yield the desired output. If operation were on gain platform C-D and an audible short duration signal occurred well above the level of B, the gain would immediately be reduced but would quickly return to C-D since the average value of the audio had not changed.

Referring once again to Figure 1, detector 3 follows amplifier V1,2 which is fed from the input. If the audio level drops below a threshold as set by the control so-labeled, detector 3 supplies gated information to the comparator and a change of control voltage is inhibited until detector 3 indicates the resumption of audio above threshold. This feature is referred to as Gated Gain Stabilizer.

In its earliest embodiments, the gain of Audimax was held constant indefinitely as long as operation remained on a given platform. As a result of operational experience, this has been modified to permit a steady gain recovery at the rate of 63% per 12 seconds. This in no way alters the interplay of the various elements described above but rather enhances the operation of the system. It has been observed that speech level of announcers tends to drop during a sentence as lung air pressure decreases. The slow recovery tends to offset this drop.

In reviewing the four requirements cited above for artistically acceptable level control, it becomes apparent that Audimax satisfies the first requirement by virtue of its fast gain reduction capability. The second requirement is satisfied by the fact that maintained changes of gain are effected by the average value of the audio. The third requirement is met in two ways; for short pauses by the finite time required for the establishment of a new average and consequent "platform", and for long pauses by the Gated Gain Stabilization described above. Finally the fourth requirement is met by providing for a return to the earlier platform following a gain reduction caused by a short audible peak. The "limiting" action for such peaks is then divided between Audimax and the transmitter peak-limiter with quick recovery on both.

Stereophonic Operation

Figure 5 is a block diagram showing the manner in which two Audimax units controlling left and right signals respectively are interconnected with the Audimax Stereophonic Adapter. The output of each Audimax is bridged and fed to a summing matrix from which a left plus right signal is derived. This is suitably amplified and applied to each of the Audimax memory units. In this manner the gain of both channels is identical and responsive to the L + R output thus preserving the stereophonic separation and perspective. Inasmuch as the modulation amplitude in FM multiplex stereophonic broadcasting is a function of the L + R signal, by ensuring the maximum amplitude of L + R we realize the optimum modulation from the system.

Operational Experience

Figure 3 is a photograph of the Audimax. It accommodates standard 19" rack mounting, is 5-1/4" high, 12-3/8" deep and weighs 27 lbs. Output level is +16 vu; -10 is minimum input for normal operation. It is recommended for use ahead of the line to the transmitter peak limiter.

Early in 1961, an Audimax was installed at

WCBS-AM in New York City and shortly thereafter on two of the network lines. Subsequently, an Audimax was placed in continuous service on WCBS-TV. During the spring of 1961, Audimax units were given field tests at four AM stations along the East coast. Audimax units are now in full-time operation at all CBS owned radio stations and at key points of the CBS television network as well as at almost one hundred independent stations across the country. In every case, a marked improvement in average modulation levels was noted without sacrifice to signal quality. Figure 4 shows two one-hour segments of WCBS-TV programming recorded on a graphic volume indicator. The lower was recorded without Audimax employing normal manual control at the originating studio. The upper chart shows similar programming with Audimax in use. The average level with Audimax exhibits a 4-6 db increase.

Conclusion

In attempting to solve the problem of artistically acceptable automatic level control, the authors have approached their task by creating a device which can simulate both the thought processes and reactions of a highly competent studio technician without many of the limitations of human control. Endorsement of the Audimax by those who are using it and the objective evidence as to its capabilities for improving average modulation have affirmed the validity of this approach.

The authors also wish to acknowledge their indebtedness to Dr. P. C. Goldmark, President of CBS Laboratories, for his enthusiastic support and to Messrs. Vorhes, Peck and Korkes of CBS Radio and to Messrs. O'Brien, Monroe and Palmquist of CBS Television for their invaluable collaboration.

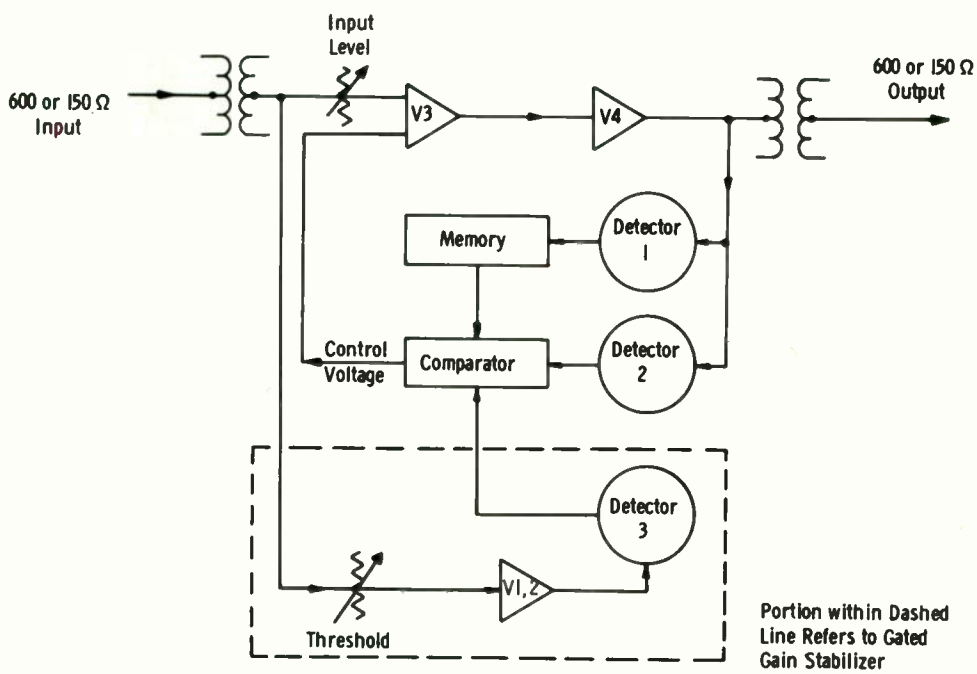


Fig. 1. System block diagram.

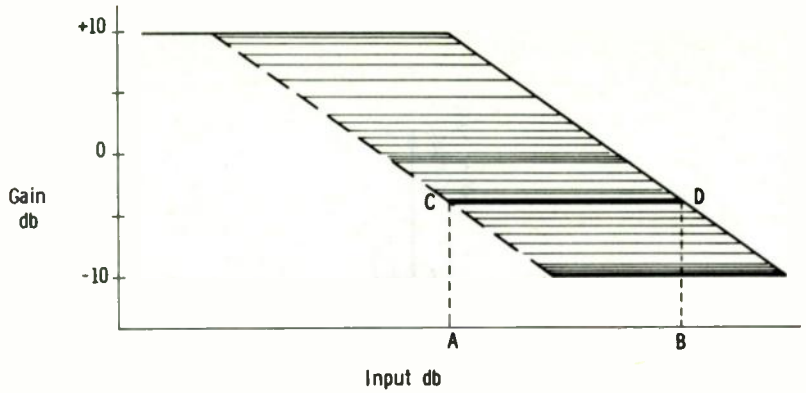


Fig. 2. Audimax performance.

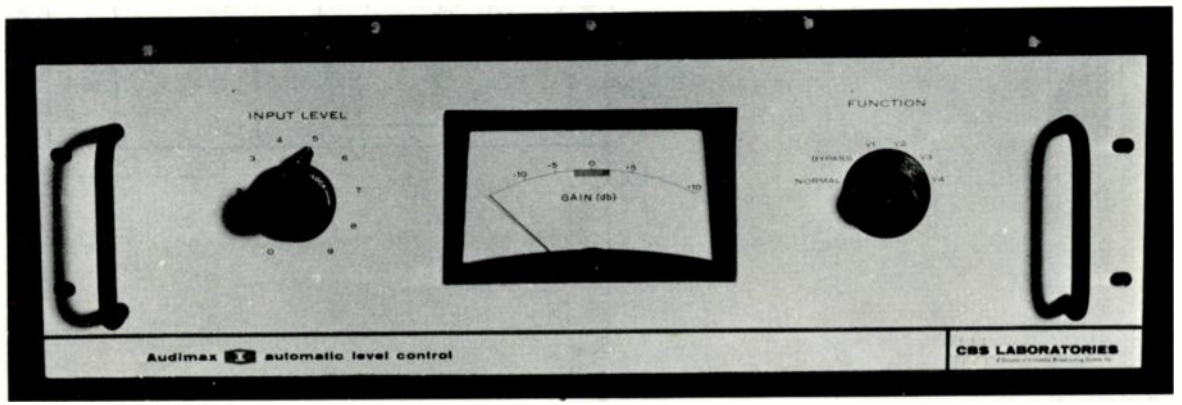


Fig. 3. Audimax.

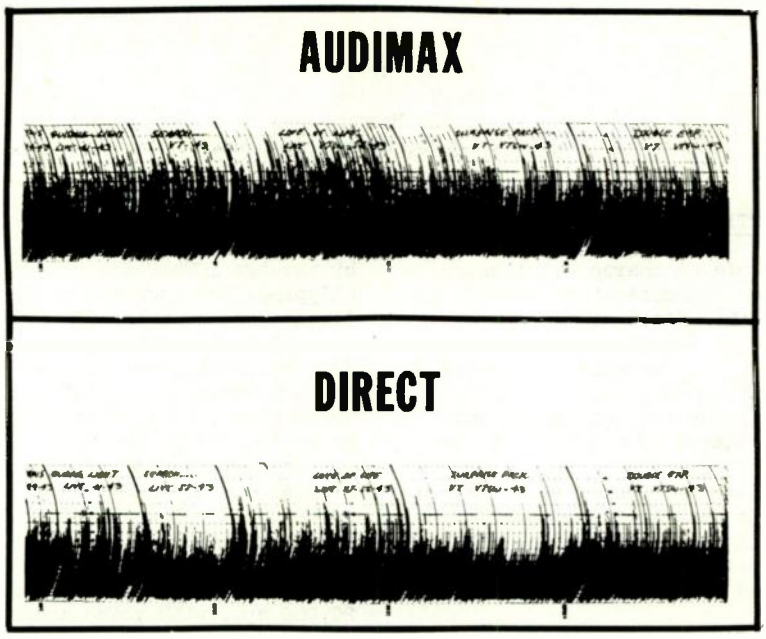


Fig. 4. Operating experience.

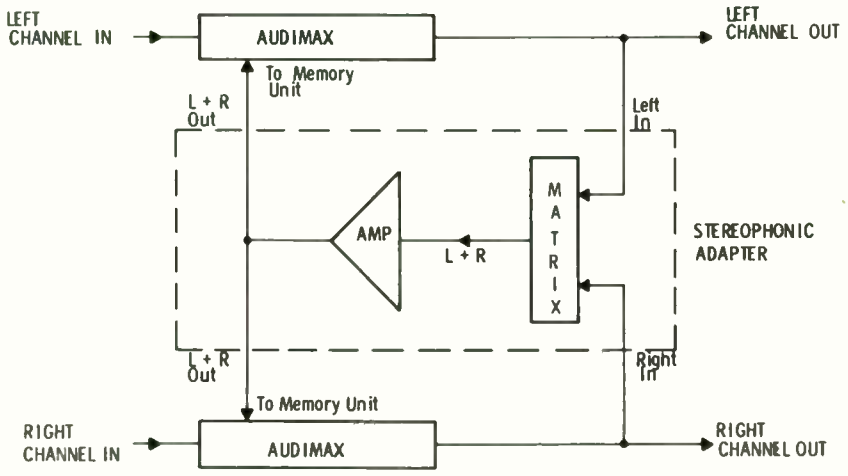


Fig. 5. Audimax stereophonic control.

STEREOPHONIC FREQUENCY TEST RECORD FOR AUTOMATIC PICKUP TESTING

A. Schwartz
G. W. Sioles
B. B. Bauer
CBS Laboratories, Stamford, Connecticut

INTRODUCTION

Since phonograph pickup characteristics are usually related to a specific test record, the recent entry into this field of the CBS STR-100 stereophonic test record is a matter of fundamental importance. The salient feature of this record is the frequency sweep bands which will yield continuous response data that can be automatically plotted by a level recorder.

The first portion of this paper will describe the test bands and their application to measurement of pickup characteristics. The second half of the paper will describe the design, calibration and mastering procedures.

APPLICATION

Pickup Response

The accuracy and ease of measurement of stereophonic pickup response frequency characteristics has been greatly improved with the introduction of the synchronized frequency sweep bands. Complete response data for left and right channels can now be obtained in the frequency range from 40-20,000 cps where heretofore only spot frequency measurements were available. Accuracy is improved by eliminating the judgement required in reading meters especially when fluctuations are present at high frequencies. In addition measurements can be made in a fraction of the time previously required.

The sweep is logarithmic at a rate of 1 decade each 24 seconds and corresponds to the chart speed of the General Radio Graphic Level recorder of 30 - 1/4" divisions per minute. The response is obtained by having the level recorder plot the pickup output to the recorded frequency sweep. A 1000 cps reference tone precedes the sweep and can be used to automatically start the recorder. Figure 1 is the schematic of such a synchronizing circuit. All relays are initially de-energized as shown in the schematic diagram. Left and right channel inputs are combined in the cathod of V1 insuring that the keying tone will be present for either direct or crosstalk measurements. The cathode follower output is fed to the V2 high gain amplifier through a high-Q LC 1000-cycle filter allowing only 1000 cps to feed through. Following this stage is a cathode follower V3 employed as a power amplifier to drive a sensitive relay K1. after the signal is rectified by the two 1N2482 diodes. A Zener diode and clipping-range control prevent high signal levels from overheating the sensitive relay.

With K1 energized, relya K2 is energized and locks itself across the power supply through contact 1. At the cessation of the keying tone, K1 is deenergized - thereby actuating K3, which starts the recorder motor at the instant the sweep begins. At the termination of the sweep band, the reset switch is manually set at RESET momentarily to deenergize K2, and the circuit is then ready for the next sweep.

A set of spot frequency bands from 20 to 20,000 cps, with voice announcements, have been recorded for manual response plotting and for playback of single tones when required. Discrete frequency bands facilitate the absolute calibration of recorded amplitudes, a procedure which will be described below.

Sensitivity

Pickup sensitivity is measured with the left and right channel 1000 cps calibration bands. The recorded level is 3.54 cm/sec rms, which is an equivalent lateral velocity of 5.0 cm/sec rms and is accurate to within $\pm$ 0.25 db.

Stylus-Tone Arm Resonance

Stylus-tone arm resonance is measured with the high level low frequency sweep bands which extend from 250 to 10 cps. These sweeps can be plotted by the level recorder and are also adaptable to automatic starting.

Compliance

Accurately calibrated 100 cps lateral and vertical modulation ranging from 1 to 5 x 10^{-3} cm peak displacement provides a dynamic method of measuring lateral and vertical pickup compliance. If the tracking force and the peak groove displacement are known, the compliance is readily calculated. This data may differ somewhat from that obtained by use of compliance meters. This is due in part to the differential force between the inner and outer groove walls arising from pickup and tone arm geometry.

Wavelength Loss

The equation (1) derived by Frank G. Miller[1] depicts the pickup response for lateral recording has been verified experimentally at CBS Laboratories and found to be applicable to single channel recording. The second factor in the equation is a function of playback frequency, pickup characteristics, and the modulus of elasticity of the disk. The first factor, in the equation a quantity that varies from 1 to 0, can be termed the wavelength loss and is a function of recorded wavelength pickup characteristics, and modulus of elasticity of the disk. For convenience in the expression shown here, recorded frequency and medium velocity i. e. diameter and recording rotational velocity have been substituted for wavelength.

The wavelength loss is measured by means of identical frequency bands ranging from 1000-20,000 cps and recorded at maximum and minimum diameters. The inner group has been compensated for the wavelength losses expected from a pickup having a 0.7×10^{-3} inch tip radius tracking at $2\text{-}4 \times 10^{-3}$ dynes. Playback of these bands under the noted condition force should produce response within 2 db between the outer and inner diameter at a given frequency. Increasing tracking force or tip radius should cause a corresponding loss in high frequency response at the inner diameter.

MASTERING AND CALIBRATION

Recording Characteristics

Selection of the recording characteristic was based on obtaining the maximum recorded level while staying within the limits imposed by the disk recording medium. Peak displacement is limited, in stereophonic recording, by the groove depth and groove pitch. At high frequencies the power handling capacity of the cutter, the Westrex 3C in this case, must not be exceeded. A peak displacement of about $.63 \times 10^{-3}$ inches and a peak velocity of 5 cm/second were selected to accommodate these limitations. The characteristic is, therefore, constant amplitude below 500 cps and constant velocity above 500 cps. The transition is abrupt, as shown in figure 2, so that variations in pickup response can be readily determined by the deviation from two straight lines. This "ideal" straight line characteristic was obtained by switching from a constant amplitude to a constant velocity characteristic at 500 cps while recording the glide tone.

Calibration

Absolute and relative calibration of the recording system proved to be the most time consuming part of the preparation. The three established methods of measuring recorded amplitudes are microscopic observation, light pattern and variable speed turntable. The first two methods, which do not require playback of the recording, are fundamental methods and are to be preferred wherever possible.

Measurement of the discrete frequency bands up to 10,000 cps was performed primarily by the microscope method. In this frequency range recorded amplitudes vary from 0.6×10^{-3} inch. Above 10,000 cps amplitudes were too small to measure by direct observation so that measurements were made directly from the projected image of a photomicrograph. Substantially the same recorded amplitudes were measured by the light pattern method. The light pattern calibration was performed by Gunter and Anderson at Shure Brothers.

Measurement of the frequency sweep bands presents a difficult problem since neither the microscope or light pattern method is applicable for continuous plotting. A modification of the variable speed turntable method was developed which was able to provide continuous calibration of the sweep bands. A block diagram of the measurement apparatus is shown in figure 3. The synchronous turntable motor is excited by a variable frequency power supply which in turn is fed by the sweep frequency oscillator that has been previously used to record the glide tone. The oscillator tuning is mechanically driven by the level recorder, and synchronous drive motor of the latter is driven by the variable frequency power supply. Playback of the sweep under these conditions will present a constant frequency to the pickup over a 20:1 frequency ratio. The characteristics of the pickup are thereby eliminated so that plotted response is that of the recording system alone.

Figure 4 is a comparison of the recording system calibration by microscope, light pattern, and variable speed turntable methods. The three measurements show good agreement. The response characteristics of the cutter head is irregular above 10KC so that the recording was made at 16 2/3 rpm at half the playback frequency. The response data shown is on a playback frequency basis.

Crosstalk measurement for either the spot frequency or sweep bands cannot be made by either microscope or light pattern because of the small amplitudes involved. Based on playback measurements, crosstalk is better than -20db from 100 to 10,000 cps.

Cutting Stylus

High frequency response of the recording system varies inversely with the size of the cutting stylus burnishing facet or "dub." To insure accuracy, the recording system must be calibrated each time a cutting stylus is installed. A damage to the stylus during the recording would require recalibration and a scrapping of the completed portions of the recording. For these reasons the cutting stylus was carefully selected so that the dub size was sufficiently small for adequate high frequency response without being unduly fragile.

Critical stylus selection was made possible by the development of a special microscope fixture. The stylus is positioned by the fixture so that it can be viewed and measured from any angle with 1300 x magnification, without being defocussed. Measurements of the dub size and tip radius are accurate to within 0.5 microns or 20×10^{-6} inches.

Mastering

Despite the increase in complexity involved, the frequency recordings were made directly from an oscillator to avoid introducing tape recorder distortion and modulation noise. Voice announcements were taped. Precise timing, which was required to synchronize the glide tones and tape recorded announcements, was achieved by use of an elaborate step-by-step procedure. Because direct recording introduces the possibility of human errors, as many as 12 lacquers of each side were cut. From this group the best two copies of each side were selected to be processed. At the processing end, careful quality control is being maintained at all stages to insure against possible deterioration in the quality of the recording.

CONCLUSION

The STR-100 Test Record has been an outgrowth of more than two years of intensive disk recording research. Among other things, this research has concerned itself with advances in calibration methods (and the development of the associated calibration apparatus), study of the cutting process and the dependence of recording response on stylus geometry, and experimental verification of Miller's pickup response function. As a result, the record is an accurate and reliable instrument with which to measure stereophonic pickup characteristics. In addition, its adaptibility to automatic testing can effect significant economics for pickup manufacturers and users.

ACKNOWLEDGEMENT

The authors wish to express their thanks to Art Gust for his valuable assistance and technical contribution during the development of the record and the work preceding it.

1 Frank G. Miller, Stylus Groove Relations in Phonograph Records, Acoustics Research Laboratories, Harvard University (Office of Naval Research TM 20: March 1950)

$$\text{RESPONSE} = \left[1-\left(\frac{f}{f_c}\right)^2\right]\left[\frac{1}{\left(1-\frac{f}{f_0}\right)^2+\epsilon^2\left(\frac{f}{f_0}\right)^2}\right]^{1/2} \quad (1)$$

$$f_c = 6.65 \times 10^{-2}\, DN\left[\frac{E}{RF}\right]^{1/3} \quad (2)$$

$$f_0 = \frac{1.95}{\sqrt{m}}\left[FRE^2\right]^{1/6} \quad (3)$$

f = frequency in cps

f_c = cut-off frequency in cps

f_0 = stylus-groove resonance in cps

D = recorded diameter in inches

N = recording revolutions per minute

E = modulus of elasticity of disk material in dynes/cm^2 (3.3×10^{10} for vinyl)

R = playback stylus tip radius in centimeters

F = tracking force in dynes

$\epsilon = \frac{r_M}{2\pi f_0 m}$

m = mass at playback stylus tip in grams

r_M = resistance at playback stylus tip in mechanical ohms

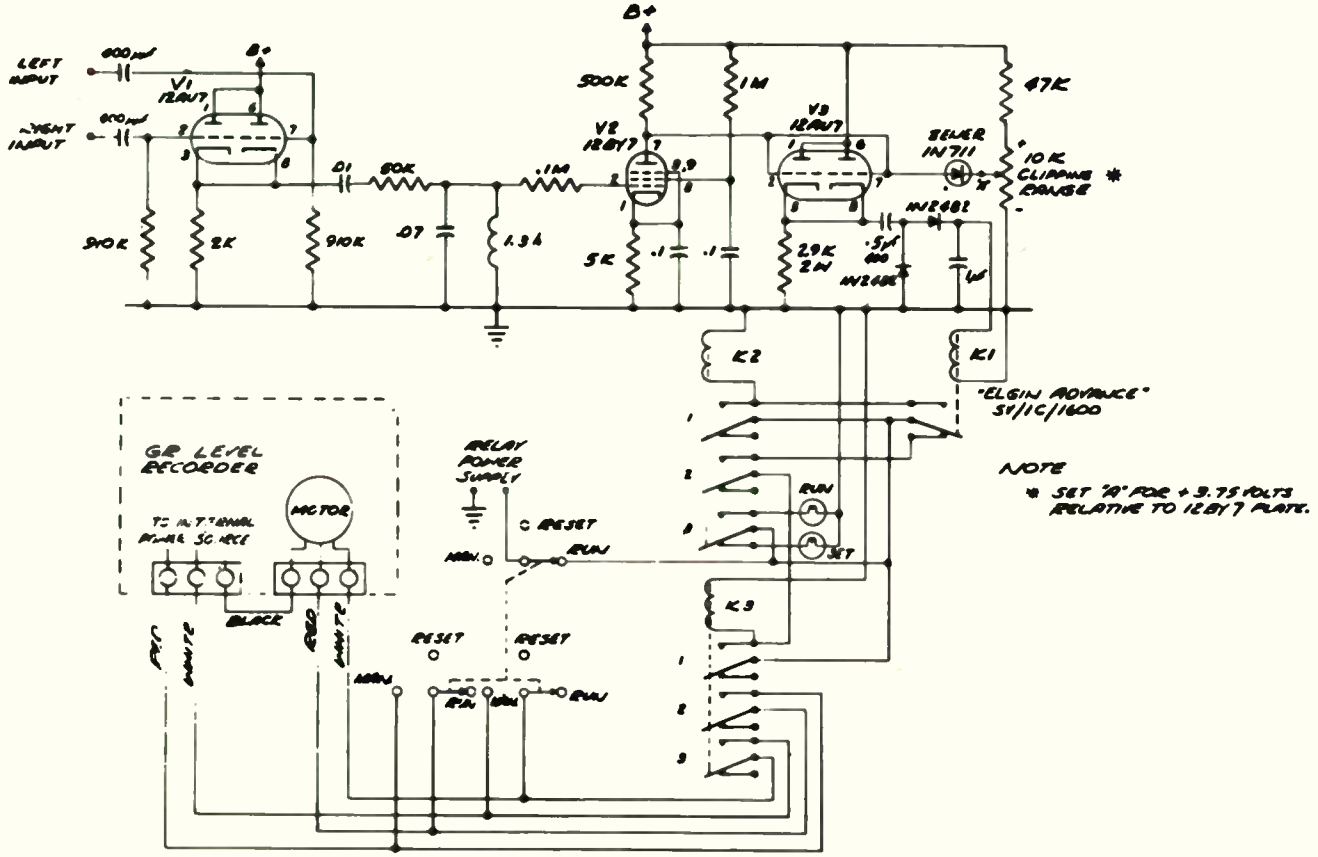
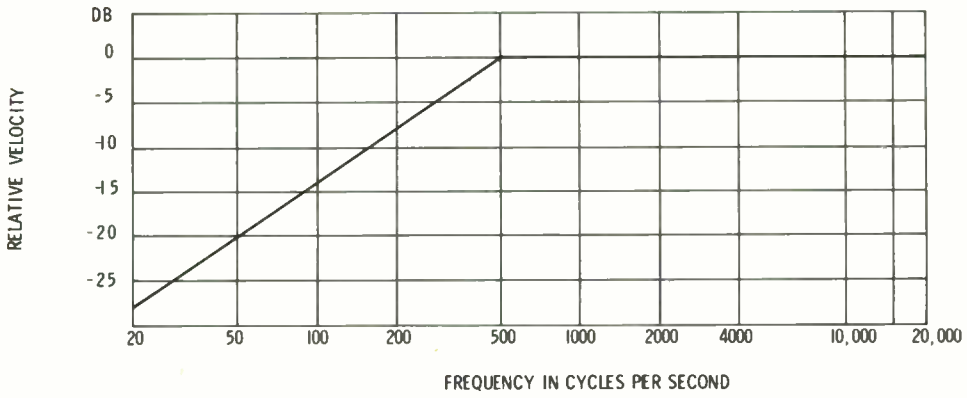

LEFT INPUT
RIGHT INPUT
V1 12AU7
B+
500K
1M
47K
V2 12BY7
V3 12AU7
SELEN IN711
10K CLIPPING RANGE *
910K
2K
910K
.07
1.34
.1
.1
5K
2.9K 2W
.5µf 600
IN2482
IN2482
1µf
K2
K1
"ELGIN ADVANCE" SY/IC/1600
NOTE
* SET "A" FOR +3.75 VOLTS RELATIVE TO 12BY7 PLATE.
GR LEVEL RECORDER
MOTOR
BLACK
RED
WHITE
RELAY POWER SUPPLY
RESET
RUN
RUN
SET
K3
RESET
RESET
RUN
RUN

Fig. 1.



STR-100 RECORDING CHARACTERISTIC

SWEEP FREQUENCY CALIBRATION APPARATUS, BLOCK DIAGRAM

Fig. 2.

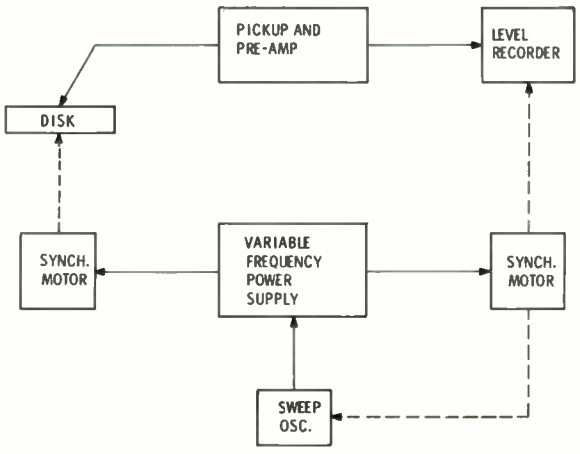
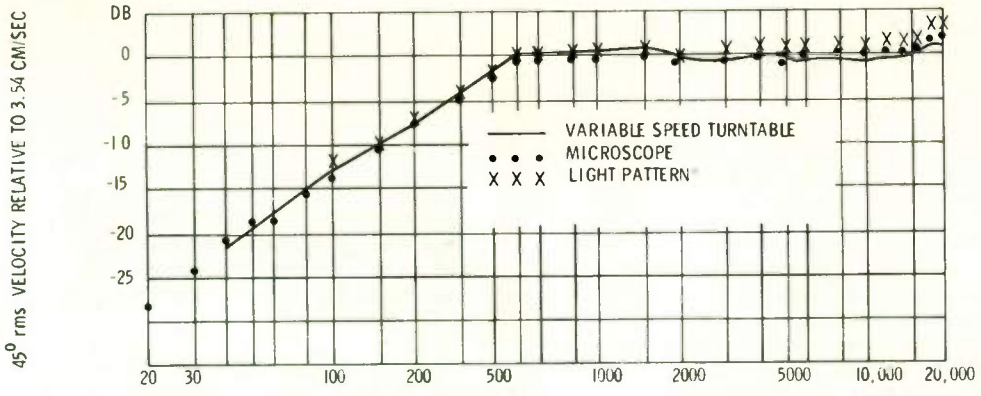


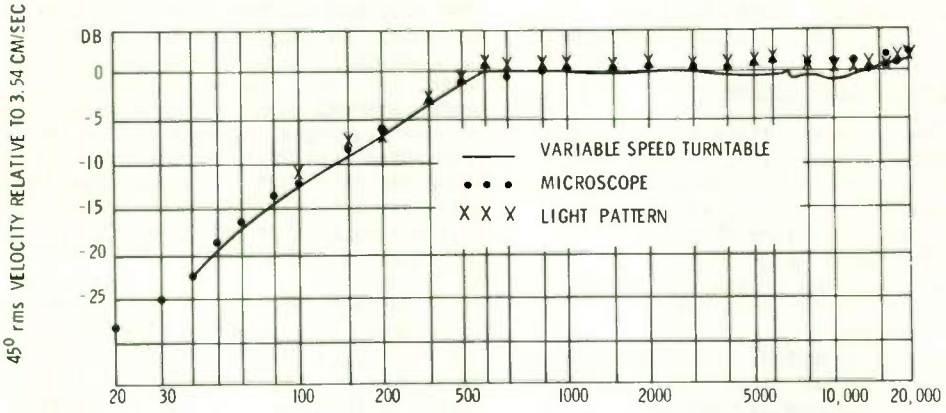
Fig. 3.



*MEASUREMENTS COURTESY GUNTER AND ANDERSON OF SHURE BROS.

STR-100 CALIBRATION, LEFT CHANNEL

Fig. 4(a).



STR-100 CALIBRATION, RIGHT CHANNEL

Fig. 4(b).

SIGNAL TO NOISE RATIO AND EQUALIZATION OF MAGNETIC TAPE RECORDING

Hanswerner Pieplow
Grundig GMBH
Nuremberg, Germany

Summary

Future tape recorders running at a speed of 1-7/8 ips must achieve an upper frequency limit of 15 kc/sec on a quarter track recording, with better signal-to-noise ratio than presently obtained with commercial recorders. This means that the subjective quality of present full track machines at 15 ips must be maintained in spite of a fifty fold increase of information density.

It is obvious that such an improvement cannot be obtained only by changes to the electronic circuitry and the like. In fact, the future requirement would be impossible to meet if improvements in tape and heads were not made. Such improvements must include new knowledge of head construction, tape-to-head contact problems, and optimum balancing of recording parameters. But once these improvements are effected, it is then necessary to check carefully if the most thorough use is being made of all available technical possibilities.

It will be shown that recording techniques have advanced to a stage where a change in the standardization characteristics is advantageous and even necessary, though they may be more complicated than the existing standards.

1. Procedure of the Investigation

Because the hum does not limit physically the recording technique and, if necessary, may be eliminated completely, we have to consider only the noise generating from the input stage of the recording amplifier, from the tape itself, and from the input stage of the replay amplifier. These noises are not correlated in any way, and each of them can be minimized without influencing the others. Since noise optimizing techniques in amplifier stages are in common use and since the absolute value of tape noise unfortunately does not depend very much on the bias level or on the constitution of the magnetic layer, there is no other chance to improve the signal-to-noise ratio except by increasing the signal on the tape.

As trivial as this conclusion looks, it does not seem to be generally appreciated. For instance, in case of essential improvements of tape and heads it is not necessary to decrease the pre-emphasis in order to match the existing standards[1], but we have to change the standards in order to increase the signal-to-noise ratio.

Thus, with the basic method of standardization in mind we have to evaluate the admittable amount of pre-emphasis and to compare the results to the actual situation of technical art.

2. Method of Standardization

Since intensity, frequency response, and distortion of a recorded signal depend upon the individual properties of the recording machine and since these properties cannot be defined in general terms, standardization can be applied only to the tape itself, i.e. to the level and to the frequency characteristic of the remanent magnetic surface flux of a recording. Once a standard tape is established, each machine can be adjusted to fit the standards.

If a magnetic tape is magnetized by a constant magnetic field, i.e. by keeping the signal current constant in the recording head, the remanent surface flux on the tape is not constant, but decreases approximately exponentially with increasing frequency (Fig. 1). Now, the basic postulate of an international standarization states that with a constant signal voltage at the input of the recording amplifier the flux vs. frequency on tape has to vary in a predetermined manner, for example, like the impedance of a resistor and a capacitor in parallel with a time constant of 200 usec.

As seen from Fig. 1, we need a certain pre-emphasis to meet this demand, coming up to 14 db at 10 kc/sec with 200 µsec and to 20 db at 10 kc/sec with 100 µsec. The smaller the value of the time constant, the higher the pre-emphasis required. Thus more signal is recorded on the tape, and a better signal-to-noise ratio is achieved.

3. The Admittable Amount of Pre-Emphasis

Of course, to avoid nonlinear distortions, a standardization cannot prescribe arbitrarily small time constants and it is necessary to

make sure that the recording system will not be overloaded. Overloading occurs theoretically when the amount of pre-emphasis exceeds in any way the slope of the spectral peak distribution of music.

Unfortunately very different spectra are published, ranging from a decrease of 18 db at 10 kc/sec[2,3] up to no decrease at all[4] in modern music, and in spite of the most ingenious considerations no reasonable method has been found to match pre-emphasis and frequency distribution of peak energy in music.

On the other hand there is no exact information about the probability of occurrence of a special event, i.e. how frequently, for example, a 10 kc/sec peak energy level will occur in all music. Therefore, following rigorously an orthodox line with regard to peak energy distribution and excluding consequently any pre-emphasis, we may run the risk to record 99% or more of all musical playing time with an inferior quality, i.e. with a smaller signal-to-noise ratio than we could do, whilst recording 1% or less of the whole musical playing time without hypothetical distortions in the upper frequency range whose audibility is not even proved.

For these reasons it looks more realistic to go back to the practical experience. In the last 10 years more than one million of tape recorders have been produced in the GRUNDIG factories with an average pre-emphasis of at least 15 db at 10 kc/sec and without any objection to distortions. This fact is a very strong argument in favour of 15 db pre-emphasis at 10 kc/sec.

4. The Actual Situation in Recording Art

Figure 1 states the average situation of head and tape, used in home tape recorders at 3-3/4 ips in 1958. A more precisely defined measurement is shown in Figure 2 and Figure 3 where the curves represent the head EMF and the remanent surface flux respectively vs. frequency at 3-3/4 ips, the head being a combined recording and reproducing half track head demonstrating the statistically mean value of a large number of heads produced in 1959, the tape being the DIN*- Bezugsband which is a German reference tape with magnetic properties guaranteed by the Physikalisch-Technische Bundesanstalt. As is to be seen from Figure 3 a 100 µsec equalization would require a pre-emphasis of 18 db at 10 kc/sec which is a fairly high value, if compared to the conditions of the measurement. This high value was the main reason why up to now the majority of German and English tape recorders were equalized following the IEC recommendations with 200 µsec instead of the NAB standards with 100 µsec.

During 1960, as a consequence of a very intense research work in the field of head design and tape manufacturing, essential improvements have been made. The curves b in Figure 2 and Figure 3 represent the actual situation and can be compared directly to the curves a with the only difference that the curves b refer to quarter track and to a tape speed of 1-7/8 ips. The very big progress issuing from Figure 2 and Figure 3 does not only refer to the level, because other factors being equal a loss of 10 db in level results from the 3 to 1 reduction of track width alone, but refers particularly to the frequency response allowing now a 100 µsec equalization at 1-7/8 ips with a pre-emphasis of 16 db. Before this was impossible or at least risky even at 3-3/4 ips[5].

5. Existing Proposals

It appears that the evolution of technical art runs faster than the standardization work of the different committees. The members of the IEC* are still arguing about 100 or 120 or 140 µsec at 3-3/4 ips. Even the last German proposal for 1-7/8 ips[6] shows a surface flux vs. frequency varying like the transfer characteristic of a network composed by two equal RC meshes, each of them with a time constant of 70 µsec. As can be seen from Figure 4, the results of such an equalization would not be too satisfying:

i) The playback amplifier needs an equalization of 210 µsec, which is too much and should be avoided with respect to a good signal-to-noise ratio.

ii) The response of an ideal playback head needs a post-emphasis of 4.5 db at 10 kc/sec.

iii) By comparison to Figure 3 the pre-emphasis whilst recording is only 5 db at 10 kc/sec. This is much too small with respect to a good signal-to-noise ratio.

6. New Equalization for Home Tape Recorders

The present state of art in Figure 2 and Figure 3 was achieved:

i) By the use of a very thin and, therefore, very pliable tape of medium coercivity. The tape must have a

*DIN German Industry Standard

*IEC International Electrical Commission

very smooth, highly polished surface as obtained for example with LGS 26 tape.

ii) By the use of separate heads for recording and reproducing, which are particularly designed to have extremely straight and extremely sharp gap edges.

iii) By carefully choosing the bias of the recording head in order to get the best compromise between level, distortions and frequency response. This bias does not coincide with that giving maximum output at 1 kc/sec. The better the tape matches high frequency performance, the smaller the difference is in bias from optimum at 1kc/sec.

It has been proved that the state of the art, demonstrated by Figure 2 and Figure 3, can be maintained in mass production. It seems that the standardization proposals existing until now have not taken this performance into account. Therefore, we have introduced a 100 µsec equalization at 1-7/8 ips in our machines and, additionally, a pre-emphasis at the lower end of the frequency range corresponding to the impedance of a resistor and a capacitor in series with a time constant of 1591.5 µsec.

The time constant of 1591.5 µsec is half the value of the actual NAB equalization which is felt to be too high because, if any pre-emphasis at all is used at the lower end of the frequency range, it should have some perceptible influence at 180 c/sec, i.e. at the third harmonic of the power frequency. Of course, this pre-emphasis has nothing to do with the physics of the magnetic recording, but it is introduced for practical reasons, because the machines become cheaper.

The complete new equalization for 1-7/8 ips, including analytical formulae of its definition is shown in Figure 5.

Referring to 3-3/4 ips, it is not favourable to change the time constant. Better signal-to-noise ratio may result in shifting the bias to the 1 kc/sec maximum point. In this way the continuity with the NAB equalization is maintained at the same time. At 7-1/2 ips the time constant has to be cut down to 50 µsec.

7. Noise Measurement

With regard to signal-to-noise ratio, it is necessary to define the signal and the noise in order to get a realistic figure of merit to judge the quality of a machine or of an equalization system.

Unfortunately, signal and noise measurements are not standardized. Of course, the signal measurement, or more precisely, the maximum signal measurement is very clear and may be achieved by measuring the output at a given frequency, say 333 or 1000 c/sec, at a given distortion, for example 5%[7]. However, it is not correct at all to measure the noise by a linear instrument independent from frequency in the whole frequency range. In addition, it is not correct to suppress the hum by a high pass filter. In order to get an objective measure of a subjective noise impression, it is necessary to take account of [8, 11] the ear sensitivity vs. frequency and vs. loudness and the spectral noise energy distribution, which is essentially "white" and which, therefore, has an increasing energy content per octave with increasing frequency. For this combination of ear sensitivity and energy distribution we hear mostly a hiss, although the hum may have a higher amplitude.

To meet the demand for the objective measurement of a subjective noise impression, the CCIF[9] has published a weighting characteristic (Figure 6) which is used by all European broadcasting stations. It is enigmatical why this widely used method has not yet officially entered the tape recording field and seems to be unknown even to professional tape recorder people[8, 10, 11]. There is no doubt that using the weighting characteristic of Figure 6 which, for example, is built into the Siemens Gerauschspannungsmesser Rel 3 U 33 the signal-to-noise ratio will become a realistic and very distinct figure of merit.

8. Noise Suppression Equalization

From Figure 6 we can deduce easily how to get the best possible signal-to-noise ratio: if the most troublesome noise frequencies are found to be between 3 kc/sec and 6 kc/sec, the signal has to be pre-emphasized mostly in this frequency region[11]. No reason can be seen why this method should not work at lower tape speeds[8], and listening tests showed clearly a subjective improvement in noise impression without appreciable overloading even at 1-7/8 ips.

A new tentative proposal for standardization of a noise suppression equalization is given in Figure 7. It has practicable values and can be defined analytically in a very closed form. If applied to normal recording and reproducing channels, it may be simply superposed (dotted lines in Figure 5).

This type of equalization looks somewhat complicated and, at the present, may not be necessary in usual home recording. But it brings up signal-to-noise ratio by 3 to 4 db and may be very useful for prerecorded music on tape of high information density.

9. Conclusions

As long as home tape recording is used just for fun or in rare hi-fi equipment, it makes no difference whether tape is wasted or not. But in case tape is used as a high quality sound carrier and has to undergo competition of disc recording, there is no doubt that high quality has to be achieved with high information densities, i.e. with small tracks and low speeds.

The broad research work which was initiated by GRUNDIG RADIO WERKE during 1960, has clearly shown that the improvements possible in the tape and head physics and technology bring tape recording into a new region no longer adequately described by the existing standards for equalization, be they NAB or DIN or IEC.

New equalization proposals were worked out and have been introduced into mass produced tape recorders with the result that, with a quarter track at 1-7/8 ips, the frequency response is flat up to 12 kc/sec within normal tolerances and that signal-to-noise ratio is better than 47 db.

An additional noise suppresion equalization could be defined which may be useful for prerecorded music.

References

1. J. G. McKnight, Journ. Audio Engng. Soc., Vol. 8, 1960 Nr. 3.
2. J. P. Overley, IRE-Transactions on Audio, 1956, Nr. 5, p. 120.
3. R. H. Snyder and J. W. Havstadt, IRE Convention Records, 1956, Part 7.
4. J. G. McKnight, Journ. Audio Engng. Soc., Vol. 7 1959, Nr. 2.
5. J. G. McKnight, Audio Engng. Soc. Preprint Nr. 171, 1960.
6. DIN 45513, December 1960.
7. DIN 45511
8. J. G. McKnight, Journ. Audio Engng. Soc., Vol. 7, 1959, Nr. 1.
9. Greenbook CCIF, Vol. 4, Geneva, 1956.
10. DIN 45510, DIN 45511
11. A. A. Goldberg and E. L. Torick, Journ. Audio Engng. Soc., Vol. 8, 1960, Nr. 1.

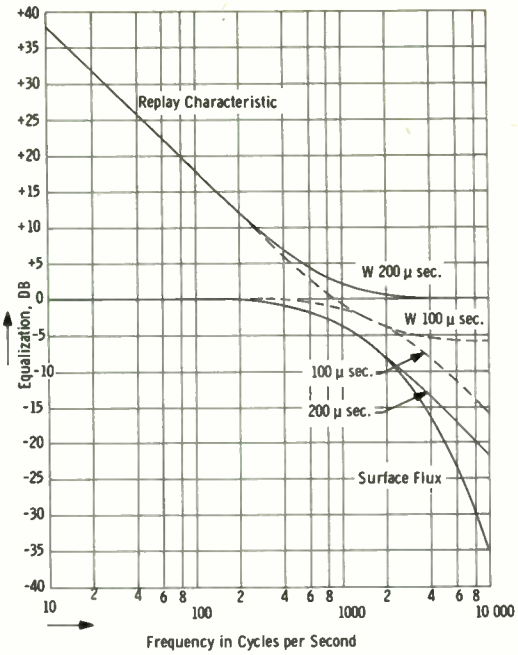


Fig. 1. Comparison of standard equalizations. 100 µsec, 200 µsec: Impedance characteristics of a resistor and capacitor in parallel with time constants of 100 µsec and 200 µsec respectively. W100 µsec, W200 µsec: Replay characteristics.

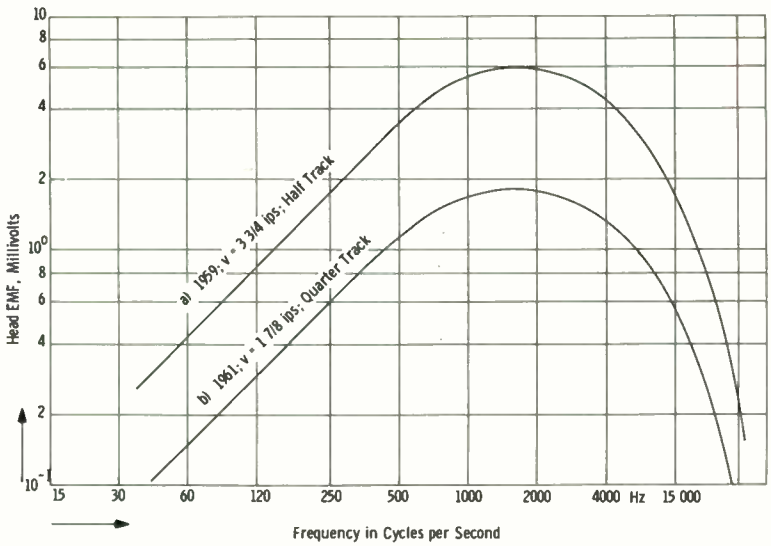


Fig. 2. Head EMF of a constant figure current recording with 4 per cent equalized distortion. a) 1959, combined half track head, Tape DIN-Bezugsband. b) 1961, separate recording and reproducing heads, quarter track, Tape LGS 26, batch Nr. 110211.

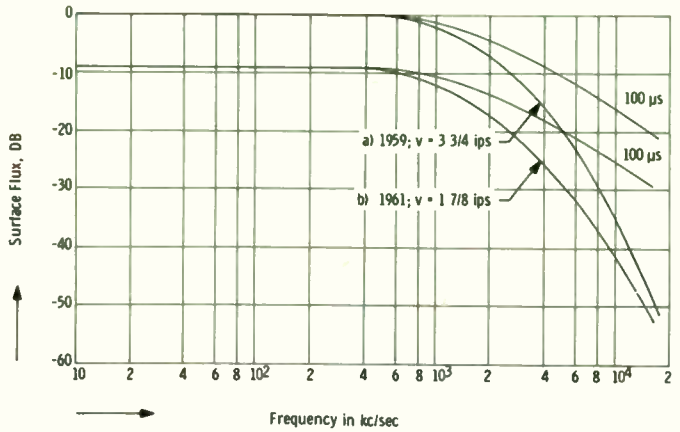


Fig. 3. Surface fluxes deduced from Fig. 2.

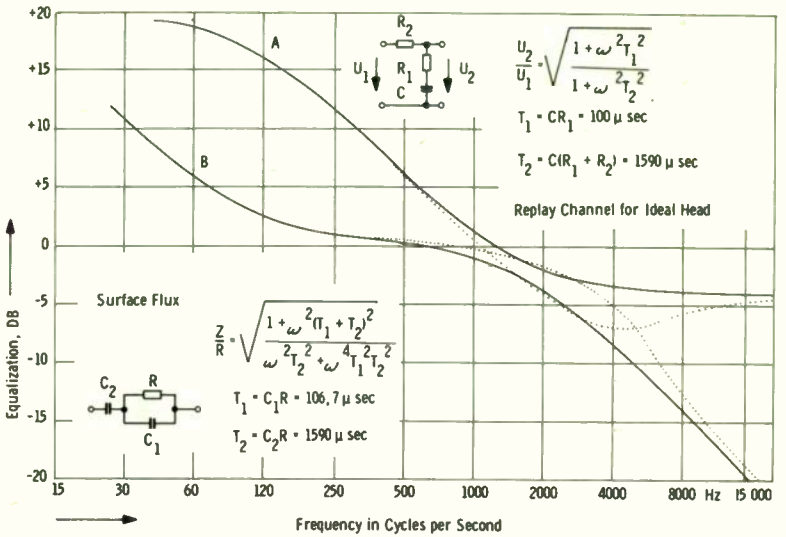


Fig. 5. GRUNDIG equalization for 1-7/8 ips (dotted lines: with additional noise suppression equalization).

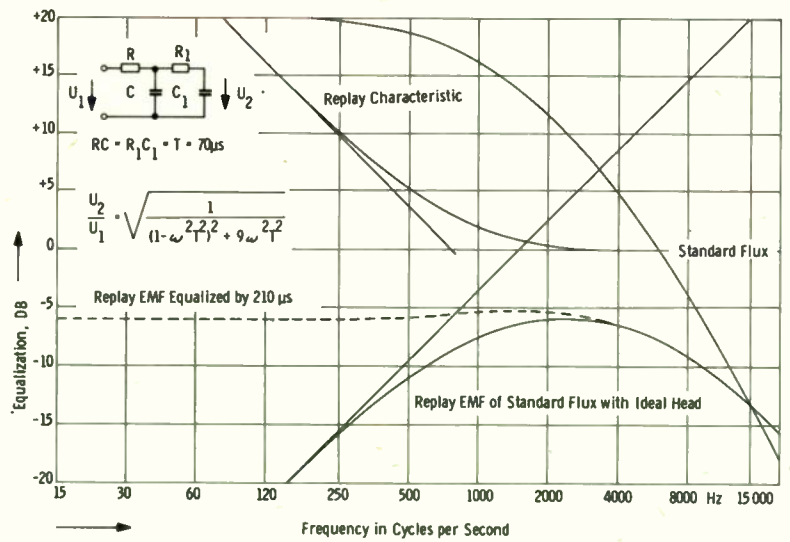


Fig. 4. DIN proposal 1961 for standard equalization at 1-7/8 ips.

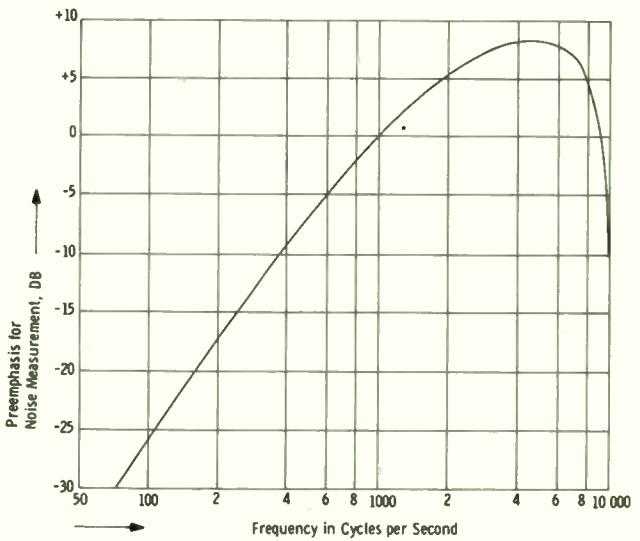


Fig. 6. Weighting characteristic for noise measurement of CCIF.

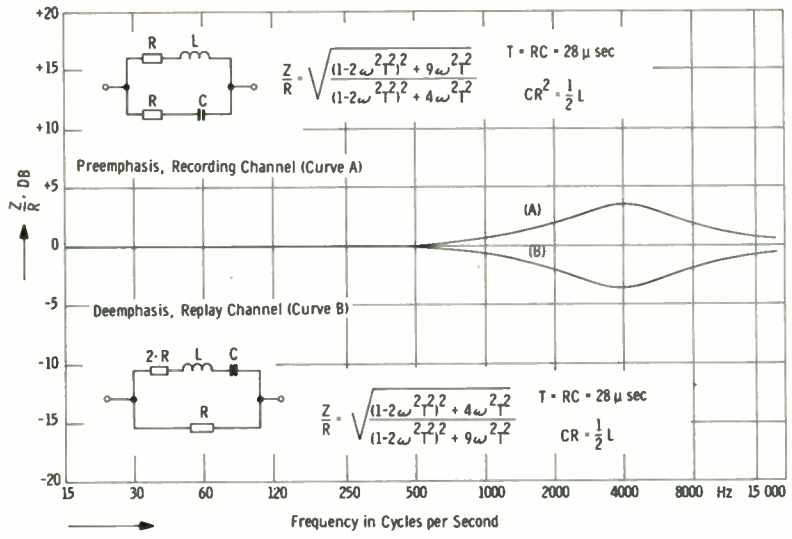


Fig. 7. Proposal of a noise suppression equalization.

NEW APPROACHES TO A-C BIASED MAGNETIC RECORDING

D. F. Eldridge* & E. D. Daniel*
Memorex Corporation
Santa Clara, California

Summary

A-c biased magnetic recording is described in a brief analysis based upon the anhysteretic magnetization process. Several different models are used to illustrate particular aspects of the recording process. It is shown that, using conventional record heads, there is bound to be a conflict between the efficient recording of short and long wavelength signals. The critical requirements involved in recording 15 kc at as low a tape speed as 1-7/8 ips are emphasized by pointing out that, on reproduction, 75 percent of the output comes from the first 0.7 micron (28 microinches) of the coating.

Various methods for improving the record process have been proposed from time to time. These methods consist of proposals for sharpening the recording field as a whole, sharpening the bias field alone, and increasing the uniformity of the recording field through the coating thickness. The design of heads to accomplish any of these objectives remains to be worked out in detail. However, several possible techniques are discussed, namely pole-shaping, eddy-current shielding and the use of poles on both sides of the tape.

An entirely new recording technique is described which utilizes successive recordings by two separate gaps of different sizes. The two gaps can be contained in separate heads or in a single head, provided that the time-delay between the two recordings, if significant, is compensated for prior to recording. The principle is simple. The long gap is fed only with low frequency signals which it can record effectively throughout the whole coating; the short gap records the high frequency signals in the outermost layer of tape. The bias field from the short gap need not significantly reduce the level of the low frequency recorded information. Filtering requirements and methods of time-delay compensation are discussed in the text, and a mathematical analysis is given of the interference to be expected from an uncompensated system.

*The work described in this paper was performed while both authors were with the Ampex Corporation Research Laboratories, Redwood City, California.

Review of Conventional A-C Biased Recording

Field from a Ring Head

The leakage field from the gap of a conventional recording head has been analyzed in some detail (Refs.1 to 6). The field distribution is complex, and there are a variety of ways in which graphical representations can be constructed. No single representation is capable of displaying all the information required for an analysis of recording. In fact, too much concentration on one representation can give rise to misleading conclusions. For example, a plot of longitudinal and perpendicular component field strengths of the type shown in Fig. 1 (b) is very useful, but only as long as it is fully realized that components are mathematically convenient rather than physically significant. From many points of view the plot of resultant field strength at various distances above the gap, shown in Fig. 1(a), is more valuable. Among other things, it shows more clearly that elements of tape passing close to the head (separation less than a third of the gap length) receive two field maxima, rather than a single maximum, these maxima occurring when the element passes above the two edges of the gap.

The way in which the direction of the resultant field changes, as an element traverses the head, is best illustrated by polar diagrams of the type shown in Fig. 2. The field is represented as a vector of length proportional to field strength and orientation relative to the direction of tape motion. The vector traces out one of the curves shown, depending upon the depth of the element within the coating. For far layers of coating, the vector traces out a circular path, the field strength being a maximum in the longitudinal direction (in the central plane of the gap) and vanishingly small in the perpendicular directions. For near layers, the curves become butterfly-shaped, the field maxima (above the gap edges) being at $\pm 45^{\circ}$ if the gap edges are perfectly sharp.

Finally, there is a third way of plotting the field distribution, which is particularly valuable in analyzing a-c biased recording. This representation consists of drawing contour lines of constant resultant field strength, as shown in Fig. 3. The contour lines are crowded in the neighborhood of the gap edges, but become more spread out further away from the gap until, at distances large compared with the gap length, the contours become semi-circles about the gap center plane.

Reproduction From Various Parts of the Coating

It is clear from the recording field plots that the tape coating will not be recorded uniformly throughout its depth c. Even if it were, the relative importance of a given layer on reproduction depends upon its position within the coating to an extent depending upon the recorded wavelength λ . If the total reproduced output is E, a layer of thickness y at the surface of the coating contributes an amount e where

$$\frac{e}{E} = \frac{1 - \exp(-2\pi y/\lambda)}{1 - \exp(-2\pi c/\lambda)}$$

This expression is readily derived from the well-known exponential law (54.5 db/wavelength) governing the loss caused by separating a recorded layer from the reproducing head. When λ is less than 1.5 c, the expression reduces to

$$e/E = 1 - \exp(-2\pi y/\lambda)$$

to within 1 percent, and the output becomes independent of coating thickness. Under these conditions, 75% of the reproduced output comes from a surface layer of thickness y_o where

$$0.75 = 1 - \exp(-2\pi y_o/\lambda)$$

or

$$y_o = 0.22\lambda$$

Taking, as an example, a 15 kc signal recorded at 1-7/8 ips, λ is 0.125 mil, and y_o is equal to 0.028 mil. In other words, under these conditions, 75% of the output comes from the first 28 microinches, or 0.7 micron, of the coating. It is instructive to compare this dimension with the average size of the oxide particles, generally about 0.7 x 0.1 micron.

Tape Magnetic Properties

The basic mechanism of a-c bias is readily explainable as an example of anhysteresis (Refs. 7 and 8). The anhysteretic magnetization process consists of applying, in addition to a small unidirectional field (analogous to the signal field), a large alternating field (analogous to the bias field) which is reduced gradually to zero before removing the unidirectional field. If the amplitude of the a-c field is sufficiently high, a plot of remanence versus d-c field strength yields a curve which is initially very linear and of high slope. In a similar way, a-c bias linearizes and increases the sensitivity of the recording process.

The most important anhysteretic magnetization curve is that showing the variation of remanence with initial a-c amplitude, using a small value of d-c field strength well within the linear range. The type of curve obtained is illustrated in Fig. 4, curve (a). The curve rises steeply as the bias amplitude approaches the coercive force. During the reduction of the bias field from a high initial strength, the remanence is acquired progressively as the amplitude of the bias field falls through a "critical range". The relative amount of remanence acquired as the bias falls by successive increments is proportional to the derivative of the curve of the remanence versus bias, implying a bell-shaped weighting curve centered approximately about the coercive force. As a first approximation, however, we can consider the remanence to be acquired uniformly as the bias amplitude falls between the values H_2 and H_1, shown in Fig. 4. In practice, the ratio H_1/H_2, when estimated by the method indicated in Fig. 4, is at best about 3/4. The mean value $\frac{1}{2}(H_1 + H_2)$ is normally approximately equal to the coercive force H_c.

The only major difference between anhysteresis as described above and a-c biased recording is that the signal and bias fields fall simultaneously, rather than separately, as an element of tape leaves the precincts of the gap. The analogous anhysteretic magnetization curve, when the fields fall together, is shown by curve (b) of Fig. 4. Instead of reaching a constant value at high bias amplitudes, the magnetization goes through a maximum at a value of bias some 20 percent greater than the coercive force.

The reason for this behavior is simply that, by the time the bias has decreased to the critical range from some high initial value, the signal field strength has also decreased in the same ratio. Consequently, using very high initial values of bias, the remanence tends to become inversely proportional to the initial bias amplitude (Ref. 8). From most points of view, the first type of anhysteretic curve is the more valuable. The discrepancy between ordinary anhysteresis and recording is not fundamental and is removed so long as we consider the effective strength of the signal field to be that existing in the region where recording takes place. It should be noted, however, that a means of recording in which the bias field falls below H_1 before the signal field appreciably decreases, would have distinct advantages over conventional recording. In particular, the bias adjustment would no longer be critical (at least as far as long wavelength recording is concerned), and some of the problems of "over-biasing" would no longer be encountered. It would be more feasible to produce a coating uniformly magnetized throughout its depth.

A final point of considerable importance concerns the variation of anhysteretic properties with the direction of the applied fields using tapes having the usual degree of particle orientation (Ref. 9). The types of result obtained are illustrated in Fig. 5. Curve (a) corresponds to magnetization along the direction of orientation (parallel to the length of the tape), and curve (b) corresponds to magnetization in a perpendicular direction. For a given signal field strength, the remanences obtained in the two directions are in the ratio of approximately 4 : 1. A very important point, however, is that the critical bias range is substantially the same in the two cases. Measurements of coercivity show that it, too, is substantially independent of the direction of orientation. One of the conclusions to be drawn from the coercivity results is that approximately the same degree of erasure should be produced when an erasing field of given internal strength* is applied in any direction.

*We shall be interested in forms of erasure that occur during recording. Intentional erasure by means of a perpendicularly applied external field is, of course, inhibited by severe demagnetization, and is consequently relatively inefficient.

Long Versus Short Wavelength Recording

For a given bias current through the head (or bias field strength deep within the gap), the decrease in bias amplitude from H_2 to H_1 occupies a certain region in space. It scarcely needs pointing out that the greatest recording resolution is achieved by making the longitudinal extent of this region as small as possible, particularly as far as the near layers of tape are concerned. Severe losses may occur when the recording region extends over a distance comparable to the wavelength of the signal to be recorded.

Now, one of the conclusions of the previous section was that the critical bias range is substantially independent of field direction. Consequently, recording should take place in the region where the resultant bias field strength falls from H_2 to H_1. Such regions can readily be determined from a contour plot of the type shown in Fig. 3. For example, if we assume a gap length twice the coating thickness, and set the bias current to a value which gives a gap field equal to H_2, recording takes place between the contour lines labeled 1 and 3/4 in Fig. 3; the recording zone is then as shown by the shaded area (1) in Fig. 6(a). This would give quite good registration of short wavelengths, but would leave a large part of the coating underbiased, resulting in severe distortion at long wavelengths. If the bias current were doubled, the recording zone would lie between the contour lines marked 1/2 and 3/8, and would shift to the shaded area labeled (2) in Fig. 6(a). Similarly, a further doubling of bias current would shift the recording zone to the shaded area (3); the whole coating would now be overbiased with loss of sensitivity at long wavelengths and very poor resolution at short wavelengths. The condition of optimum long wavelength registration corresponds approximately to setting the bias current so that the farthest layer of the coating just receives the bias field amplitude H_2. This condition is more or less satisfied by recording in zone (2) of Fig. 6(a). A somewhat lower value of bias would be needed to achieve a reasonable compromise between the two desired requirements of long wavelength linearity and good short wavelength resolution.

The situation using a small record gap, equal to two fifths of the coating thickness, is depicted in Fig. 6(b). When the bias current is set to a suitably small value corresponding to zone (1), resolution in the near layers is

greatly improved , but the effective field extends into only about the first one-tenth of the coating, and the major part of the coating is grossly underbiased. In order to obtain satisfactory long wavelength recording, the bias must be increased so that the recording zone moves to position (2). Under this condition, the short wavelength resolution would, if anything, be worse than that obtained using the optimally biased long gap (Fig. 6(a), zone (2)).

To summarize, it is impossible in a conventional system to obtain optimum short and long wavelength recording at the same time. The best that can be done is to choose those values of the ratio of gap length to coating thickness and the bias current that give the best compromise over the range of wavelengths of interest.

A further point is that under all the conditions depicted in Fig. 6, recording takes place in zones where the field is predominantly perpendicular rather than longitudinal. Actually, the field will not turn very much past 45° but, even so, it is clear from the anhysteretic curves of Fig. 5 that full advantage is not taken of the high sensitivity in the direction of particle orientation. What may happen is that an efficiently recorded longitudinal signal may, crudely speaking, be erased by a subsequent rotation of the bias field towards the perpendicular direction without adding any appreciable perpendicular component of recorded magnetization.

Improved Recording Methods

Possible Techniques

At first sight, there are three approaches, any one of which may lead to a considerable improvement in the short-wavelength recording performance. The first and perhaps most obvious technique is to "sharpen" the field in the zone of recording. In this case, both the bias and signal decay equally as the element of tape leaves the recording area. The second approach is to reduce the bias separately from the signal and to reduce it first. That is, the bias must be reduced to a value below which it can no longer influence recording, while the signal strength remains at, or near, its maximum value. This would eliminate the over-biasing problem entirely, and would, therefore, provide a system wherein the same bias current setting would be uncritical and equally good for all wavelengths. The third approach is to produce a more uniform bias throughout the thickness of the oxide. If this could be accomplished, a given bias setting would be equally good for any point within the depth of the oxide. Short wavelengths, which are recorded at the surface, would not be over-biased.

Unfortunately, these ideas are very difficult to put into practice. Reducing the length of the trailing pole of the record head has been suggested as a means of sharpening the field. This can, however, have only a relatively small effect, for the record flux will by no means limit itself to the shortened pole surface, but will spread round it to enter other parts of the pole. Westmijze (Ref. 10) has proposed limiting this "spread" by eddy-current shielding which, since the shielding would be most effective at the higher bias frequency, puts this proposal into the second category. The problem here is that it is extremely difficult to engineer a head of this type with the small dimensions of pole and conducting shield demanded by a practical system. As far as improving field uniformity is concerned, extending the head structure to the other side of the tape is an obvious technique, although one which is unattractive from an operational point of view. Three pole heads of this type have been suggested by Camras (Ref. 11), but the optimal dimensions, magnetic potentials and disposition of the poles are critical and difficult to determine.

There is, however, a fourth approach to resolving the long-wavelength linearity versus short-wavelength response dilemma which is essentially very simple. That is to record the long and short wavelength parts of the signal sequentially, using heads of different gap lengths. This technique is discussed in detail below.

Multi-Gap Recording

First, the relatively long-wavelength information is recorded by a head having a relatively large gap, capable of producing a uniform bias and signal field throughout the oxide and hence a good low frequency performance. At some later time, the short-wavelength information is recorded, using a very small gap, in which the bias is adjusted to a low value for optimum recording of very short wavelengths. Because the field drops off very rapidly with distance away from the head surface, and because the bias is adjusted to a low value for optimum recording of short wavelengths, only a very small portion of the long wavelength information is erased during this

second recording. The fields in the two heads are illustrated in Fig. 7. It would be possible to use the same head, or heads with the same gap dimension, but with the bias adjusted to different values for the long wavelength and short wavelength recordings. This method would provide a considerable improvement over the ordinary method, but would not be as effective as having gaps of different sizes.

If no correlation were required between the short wavelength and long wavelength information to be recorded, we could record these two signals at entirely different times and places. For most applications, however, the long and short wavelengths are parts of a complex signal and must, therefore, be recorded in synchronism with each other. This problem may be handled in several different ways. Either the distance between the points at which the two recordings are made must be very small, so that the delay is negligible for the particular type of recording being made; or a compensating delay must be introduced in the high frequency portion of the signal to be recorded by the second head. The first may be accomplished either by placing two record heads extremely close together, or better still, by constructing a single head which has two gaps and in which sufficient isolation between the two gaps is obtained. Such a head is shown in Fig. 8. It is possible to achieve good isolation with this type of head, although interconnected windings may be necessary at various locations on all three legs.

A compensating delay may be provided by various means; ultrasonic delay lines are available, with delays up to quite a number of milliseconds, which might be usable for this application. Another form of delay is the use of a drum recorder, where the information is recorded at one point and picked off any desired degree of rotation later on. For the making of duplicate tapes, a very simple technique may be used to provide the required delay; that is, two separate reproduce heads may be placed on the machine on which the master is being played. The first reproduce head would be used to provide the low frequency information for the first record head, and the second reproduce head would provide the high frequency, short wavelength signal for the second record head. The reproduce heads would, or could be, identical and would merely be fed into separate reproduce amplifiers, filters, and record amplifiers.

With any method of delay it will generally be desirable to pass the low frequency and high frequency portions of the signal through separate low pass and high pass filters respectively. Since the bias on the second head is adjusted optimally for the recording of very short wavelengths, any long wavelengths recorded by this head would be distorted to some degree. The overall response contributed by the two portions of the system is shown in Fig. 9(b) and (c). The overall output will simply be the sum of these two transfer characteristics. It is obvious that there will be some interference between the signals at the cross-over point. If the two are out of phase, there will be some degree of cancellation; and if they are in phase, there will be addition. This may be analyzed as follows: -

Let the double-gap recording system have the characteristics illustrated diagrammatically in Fig. 9; Fig. 9(a) indicates the way in which recording takes place at the two gaps. The separation "d" between the two recording zones leads to a time difference $T = d/v$ between the recorded signals. Fig. 9(b) shows the response-frequency characteristics of the long-gap recording channel, including the effect of the bias from the short gap plus low-pass filtering and replay equalization. The transfer function of this essentially low-pass characteristic is represented by

$$A = K\, g(f),$$

where $g(f)$ is a decreasing function of f of initial value $g(o) = 1$. Fig. 9(c) shows the corresponding characteristic of the short-gap recording replay channel. If the response characteristic of the double-gap system as a whole is to be flat, the transfer function of the short-gap channel is given by

$$B = K\,(1 - g(f)\,).$$

Recording with both gaps simultaneously, we get an output proportional to

$$A \sin \omega t + B \sin \omega\,(t - T)$$

$$= \sin \omega t\,(A + B \cos \omega T) - \cos \omega t\,(B \sin \omega T),$$

so that the transfer function of the combined channels is given by C, where

$$C^2 = (A + B\cos wT)^2 + B^2 \sin^2 wT$$

$$= A^2 + B^2 + 2AB\cos wT$$

$$= (A+B)^2 - 2AB \sin^2 (wT/2)$$

$$= K^2 \left[1 - 2g(f)\,(1 - g(f)) \sin^2 fT \right],$$

or the normalized transfer function is given by

$$C/K = \left[1 - 2g(f)\,(1 - g(f)) \sin^2 fT \right]^{1/2}.$$

Minima in the response will therefore occur when

$$\pi fT = \pi/2,\ 3\pi/2,\ 5\pi/2,\ \cdots\ (2n+1)\pi/2,$$

i. e. at frequencies given by

$$f_o = 1/2T,\ f_1 = 3f_o/2,\ f_2 = 5f_o/2,$$

$$f_n = (2n+1)\, f_o/2\ ,$$

and the ratio of the output of the n^{th} minimum to the maximum output will be given by

$$\left[1 - 2g(f_n)\,(1 - g(f_n)) \right]^{1/2}$$

and has its least value of 1/4, or - 12db, when $g(f_n) = 1/2$, that is to say, at the cross-over frequency f_c.

Typical values for a practical system are:

d = 6 mil

v = 1.85 ips

T = 3.2 milliseconds

f_o = 156 cps

Minima will therefore be spaced 156 cps apart in the region of the cross-over frequency (normally about 1.5 kc)and have a maximum possible depth of 12db. How many significant minima there are, and how nearly the worst one approaches a depth of 12db, will depend upon the sharpness of the filtering and the choice of the cross-over frequency. Ideally, if the cross-over frequency is adjusted to be an integral number of f_o's, and the filters effectively cut off whithin a frequency range $f_o/2$, there will be no minima. This argument, of course, ignores any of the miscellaneous phase shifts which must inevitably exist in the two channels, and in practice the required filter adjustments will tend to be critical.

A preliminary experimental verification of the multi-gap technique was obtained at 1-7/8 ips with the results shown in Figs. 10(a) and (b). Long-wavelength recordings were made, using an ordinary head having a 0.5 mil gap with the bias adjusted to obtain maximum output at 125 cps (14db below saturation level). Short wavelength recordings were made using a head having a 0.1 mil gap with the bias adjusted for maximum output at 10 kc. In each case the record current was adjusted to give an output 14db below saturation at the frequency used for setting the bias, and this current was used throughout the subsequent tests. The partial erasure of the long wavelength recordings by passing them over the 0.1 mil gap, energized by bias alone, is shown in Fig. 10(a). The erasure is less than 2 db for frequencies below the probable cross-over frequency of approximately 2 kc. The frequency response of the 0.1 mil gap recording is plotted relative to the response of the 0.5 mil gap recording in Fig. 10(b). It seems from the curve that the potential improvement to be gained from a double-gap technique incorporating the above gap dimensions and bias conditions is approximately 12db/octave. Extrapolation to 15 kc indicates an improvement of some 36db relative to the admittedly somewhat overbiased 0.5 mil gap recording.

Conclusions

The multi-gap recording technique has proved to be effective in overcoming the frequency-response limitations of conventional a-c biased recording. In fact, this technique, or a modified form of it, may well prove to be the best for many applications. It is certainly the simplest. Initial tests showed that the potential gain in 15 kc response at 1-7/8 ips was more than 30 db using conventional tape; these results have been confirmed in later and more practical forms of the device. Time delay compensation is not too serious a problem provided duplication - rather than direct - recording is involved, because correction can be introduced by reproducing the master tape using two suitably spaced gaps. Flutter requirements may be critical when compensating for long time delays but should be within the range presently achieved on high quality tape transports.

It is probable that a further major improvement must await the production of more suitable tapes. Use of the multi-gap recording technique enables the audio-frequency range to be covered adequately at 1-7/8 ips, but the back-

ground noise after equalization is relatively high. A better tape would have a lower noise level in the presence of bias, or require less equalization, or would incorporate both these qualities.

Finally, it should be mentioned that the use of the multi-gap recording method need not be confined to low speed audio recording. It could be used to improve the performance of higher-speed, audio-frequency recording systems and also analog systems going to higher frequencies. In particular, it could very easily be used to increase the number of channels in a telemetry data recorder. The problem of time-delay compensation need not arise in the latter case, if the channels are suitably divided between the heads (or gaps).

References

1. A. D. Booth, Brit. J. Appl. Phys. 3 307 (1952).
2. W. K. Westmijze, Philips Res. Rep. 8 148 (1953).
3. O. Karlqvist, Kungl. Tekniska Hogskolans Handlingar, #86 (1954).
4. J. Greiner, Nachrichtentechnik 5 295 (1955) and 6 63 (1956).
5. G. Schwantke, Acustica 7 363 (1957).
6. S. Duinker, Tijdschrift Van Het Nederlands Radiogenootschap, 22 29 (1957).
7. W. K. Westmijze, Philips Res. Rep. 8 245 (1953).
8. E. D. Daniel & I. Levine, J. Acoust. Soc. Amer. 32 1 (1960) & 32 258 (1960).
9. R. F. Dubbe, IRE Trans. (Audio) AU-7, #3, 76 (1959).
10. W. K. Westmijze, U. S. Patent 2,854,524 (1952).
11. M. Camras, Armour Research Foundation Bulletin #73, (September 1951).

Added Note: -

Since the writing of this paper, a US Patent has been issued to Mr. Derk Kleis, Eindhoven/Netherlands, assigned to the Philips Company, which describes a dual gap recording system which is very nearly identical to that discussed herein. The work at Philips was unknown to the authors at the time of the original preparation of this paper in 1959.

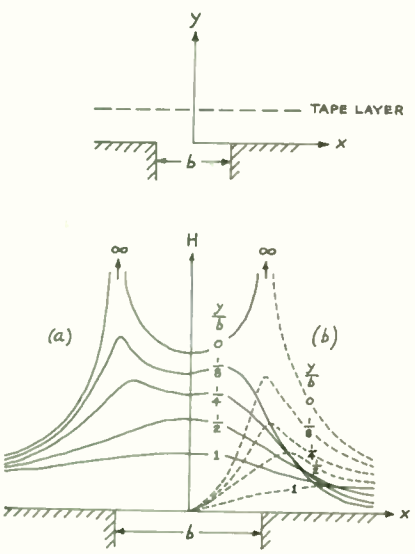


Fig. 1. Distribution of recording field. a) Resultant field. b) Longitudinal (solid) and perpendicular (dashed) components.

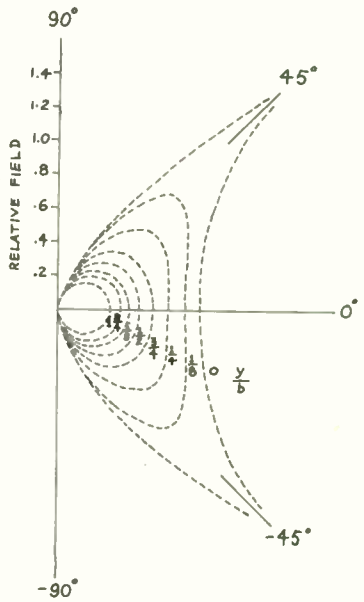


Fig. 2. Loci of resultant field vector applied to various layers of tape.

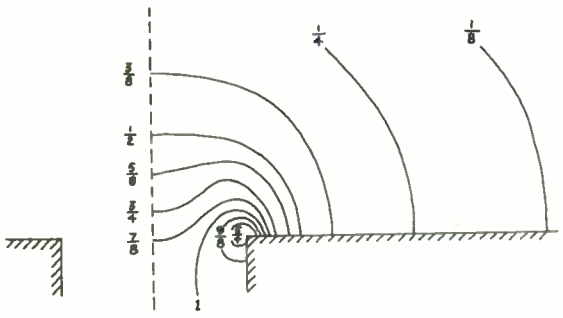


Fig. 3. Contours of equal resultant field strength, relative to the strength deep within the gap.

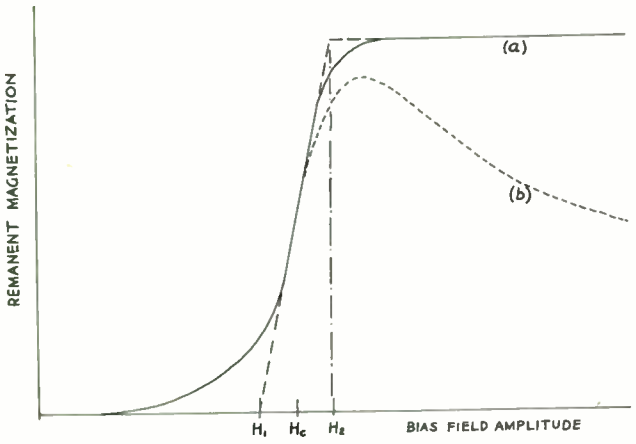


Fig. 4. Anhysteretic magnetization curves for a small signal field. a) Bias reduced to zero before signal. b) Bias and signal reduced to zero together.

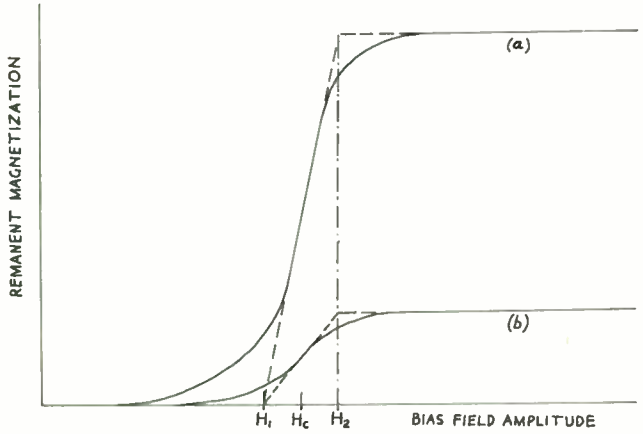


Fig. 5. Anhysteretic magnetization curves for a small signal field. a) Bias and signal fields parallel to tape length. b) Bias and signal fields perpendicular to tape length.

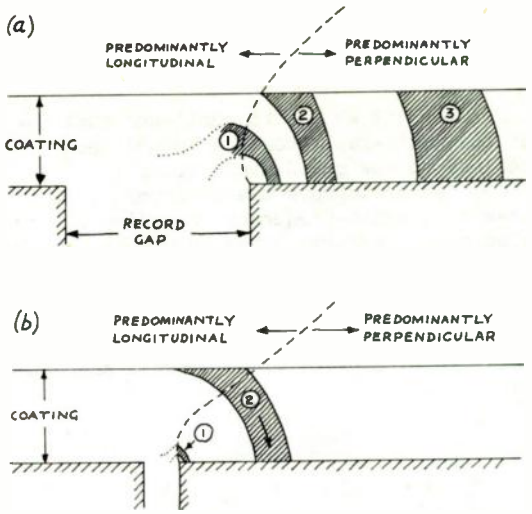


Fig. 6. Recording zones. a) Large gap. b) Narrow gap.

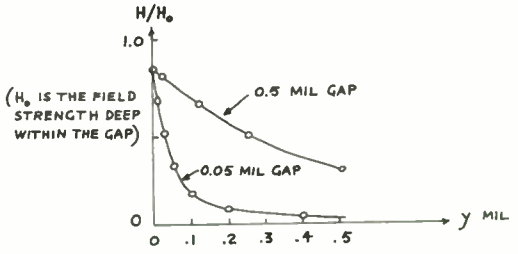


Fig. 7. Fields at gap centerplane from heads with broad and narrow gaps.

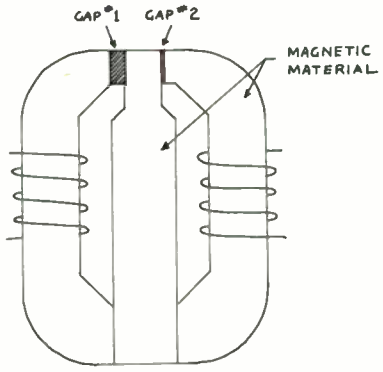


Fig. 8. Two-gap record head.

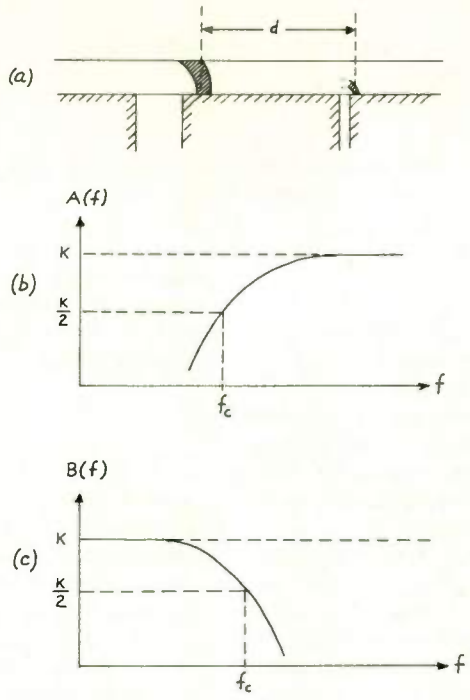


Fig. 9. Illustration of two-gap record process. a) Recording zones. b) High-pass characteristic (short gap channel). c) Low-pass characteristic (long gap channel).

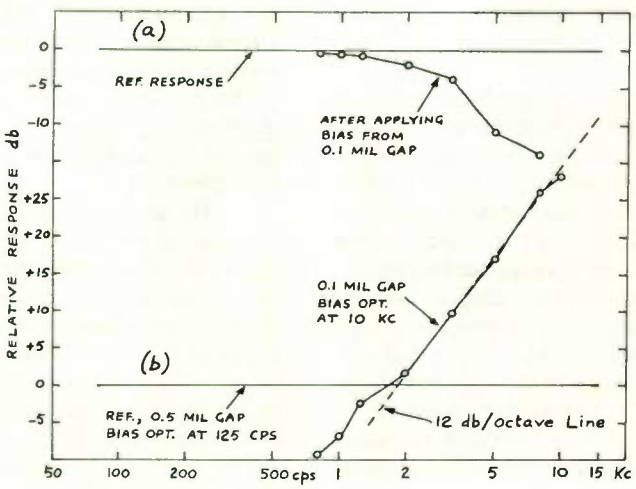


Fig. 10. Response-frequency curves. a) Partial erasure of 0.5 mil gap recording by 0.1 mil gap bias. b) Comparison of short gap and long gap response.

ANALYSIS OF TAPE NOISE

Irving Stein

Research Division, Ampex Corporation
Redwood City, Calif.

Summary

Noise and the signal-to-noise ratio arising from the transduction of a storage medium such as a magnetic tape is analyzed for the case where the storage elements (magnetic domains) are non-interacting and the magnetic tape, except where otherwise specified, is homogeneous. Since the fields arising from the tape are uniquely determined only if the boundary conditions (established by the presence of the reproduce head) are given, the broad-band signal-to-noise ratio is a function not of the tape alone, but of the tape-head system. It is found that the signal-to-noise ratio is proportional to the square root of the 'effective' number of particles detected by the head. Thus, the signal-to-noise ratio is proportional to the square root of a length of tape, which is a measure of the tape head-resolution. The noise itself can be considered to arise from sources which add, in circuit terminology, serieswise both across the width and along the length of the tape.

The essential source of the no-signal tape noise is not in the surface roughness (assuming an otherwise homogeneous tape) or the physical distribution of the particles, but in the random orientation of the magnetic moments. Furthermore, a distribution in either the particle magnetization or the particle volume will increase the noise. These conclusions are derived from purely statistical considerations.

Although the basic concepts of tape-head noise can be established without considering interaction of the tape particles, the formulas for d.c. and a.c. signal-to-noise and correct quantitative results for the no-signal noise can only be obtained if the effect of particle interaction is included in the analysis. An externally applied field divides the particles into two groups:'unflipped' particles, which strongly interact, and 'flipped' particles, which weakly interact. Using this distinction, it is possible to derive the signal-to-noise ratio for d.c and a.c. magnetized tape with no further reference to particle interactions. The formula for signal-to-d.c. noise is found; for small magnetizations it is seen that the noise voltage is proportional to the square root of the magnetization level. A.C. noise is found to be purely a modulation of d.c noise. Obtaining quantitative results, of course, involves interaction theory, a subject not treated here.

The author wishes to thank Paul Vadapalas of the Ampex Research Laboratory for plotting the curves in Figures 2 and 5.

1. Introduction

In all physical systems, the storage or transmission of a signal is the storage or transmission of a phenomenon averaged over a small region in space or time. Due to the inherent variability of nature, at any instant of time-space the signal varies from its average. In many physical systems this variation is called 'noise'. Usually noise refers to a time fluctuation in a signal. In this paper we discuss noise arising from a space variation in the signal such as would occur in a storage medium, e.g. a magnetic tape, as it is transduced.

A magnetic tape contains a multitude of small magnets ($> 0.5\mu$) or 'domains', which we assume are randomly dispersed and oriented. Any given signal or bit of information is recorded on a host of such magnets, a collection large enough to form a statistical ensemble no matter how small the bit. Thus, a signal is constant over a region of tape if its average over a statistical ensemble is constant over a region of tape. However, at each particle or domain within the statistical ensemble the signal will vary. This random variation is the source of noise. It is necessary to point out that only random variation and not discrete variation of magnetic sources produces noise. Thus, a medium with continuously varying random magnetic sources is also a noise source. (Continuity, of course, does decrease randomness generally.)

The purpose of this paper is to establish basic concepts in the theory of magnetic tape noise and from these concepts determine the functional relationships among the relevant parameters. The approach is based upon a simple statistical analysis of the magnetic domains in the tape. Since the noise, and the signal-to-noise ratio, is detected by a reproduce head, it is necessary first to evaluate the tape-head system and then determine the degree to which tape and reproduce head can be considered separately in the evaluation of noise. Background, d.c., and a.c. noise will be analyzed using the same approach.

A general analysis of noise can be considered as consisting of two phases: (1) an analysis based upon non-interacting domains; and (2) an analysis based upon interacting domains. Only the latter phase can secure accurate quantitative results; however, this approach will inevitably expand to include other concepts as well as noise, such as coercivity, remanence, anhysteretic susceptibility, and ultimate tape information capacity. Therefore, since the basic concepts of noise are established by phase (1), we deal primarily with it and consider only the results of phase (2) analysis which are fundamental to an explanation of d.c and a.c. noise. In this paper, we do not necessarily expect accurate quantitative results from an analysis of noise resulting from actual tapes, which generally consist of interacting domains, even though the tapes be homogeneous. We do expect, however, the functional relationships among the relevant parameters to be true, even for interacting domains, except where otherwise indicated.

2. The Physical System

If a magnetic tape with a signal on it is passed by a transducer susceptible to magnetic fields (magnetic head, conducting loop, Hall effect head, etc.), an output signal is produced by the transducer, generally in the form of a voltage. Both the magnitude and the accuracy of the response is determined by the geometry of the transducer, the tape transducer spacing, and the inherent material characteristics of the transducer, such as its frequency response, susceptibility, etc. The small variations in the signal on the tape due to the random magnetic sources in the tape are themselves detected as small random signals (noise) by the transducer. The impressed signal and the noise are detected by the head in a similar, but not identical, manner. Therefore, an attempt to 'normalize' the noise, i.e., to eliminate the effect of the head (and the strength of the magnetic source) by taking a signal-to-noise ratio (S/N), is only partially successful. Thus, in analyzing noise, and also S/N, it is necessary to consider the entire tape-head system and not just the tape itself, including tape-head spacing, head geometry, and other system factors. The factors which are relevant are then determined by the analysis. It will be found that S/N is a function of all these factors, except that narrow band S/N, i.e. the ratio of signal and noise of the same wavelength, is obviously a function of the tape only and not of the head.

Particular attention must be paid to a description of the tape itself. The magnetic layer on the tape is a thin layer ($\sim$ 0.4 mil) of magnetic sources. These magnetic sources are small particles generally assumed to be single domain particles. A single domain is defined as a permanent magnetic dipole whose magnetic vector direction, but not its size or moment, is a function of the applied field. In actual magnetic tapes there is much clumping of particles. If domain nucleation exists at the points of contact, then it is possible that two or more particles would form a single domain. However, it is assumed here that practically all of the particles are small enough to be isolated single domains and large enough (out of the superparamagnetic region) to be permanent magnet dipoles. The particles themselves can be acicular ($< 0.5\mu$ x $< 0.1\,\mu$) or more uniform in shape, such as cubic or hexagonal ($< 0.1\,\mu$).

It is to be noted that the main limit to the recording of very short wavelengths is the recording resolution, which is determined by the gap width and tape-head spacings, and by the demagnetization within the tape. However, if the recording resolution were perfect and we disregarded demagnetization, then wavelengths considerably smaller than a particle length could be recorded. This is mentioned only to emphasize the fact that noise wavelengths much shorter than particle lengths also exist on the tape. However, since the bandwidth ($<$ 1 mil) contains only a small portion of the total noise, we can assume all wavelengths subsume a statistical ensemble of particles.

We assume we are investigating single domain particles (described by the Stoner-Wohlfarth Theory[1]), randomly distributed with no interactions. (Only in the particles of d.c and a.c. noise will the effect of particle interactions be introduced.) The particles are assumed to be less than 1μ in their longest dimension and, on the average, less than 1μ apart. As the tape passes by the reproduce head, some of the flux from each magnetic particle enters the head. Since the particles are assumed to be randomly oriented (completely demagnetized tape), the flux entering the head varies randomly with the tape motion. If there is a signal on the tape, the flux entering the head varies randomly, to a small degree, around the signal level. The degree of variation determines the noise of the head-tape system and occurs with or without a signal present on the tape. It is this noise which is the subject of investigation in this paper.

3. Statistics of Noise

A distinction is to be made among the terms 'magnetic source', 'noise source', and 'noise'. The 'magnetic source' is the magnetism of the particles. The 'noise source' is the variation of the magnetization from its mean. For no-signal noise, the mean is zero. The noise source is defined more specifically as the standard deviation of the magnetic moment, $\sigma(m)^*$, where m is that component of the particle moment giving rise to the head-detected field. On the other hand, 'noise',

* All notation is defined on page 17.

N, depends not only on $\sigma(m)$ but also on other factors yet to be determined. Thus, noise arises from the noise source, and the noise source arises from the magnetic source. Since the tape is assumed to be homogeneous and the tape velocity is constant, the noise per unit velocity of an equalized head is given by a direct calculation of the noise flux. The system is shown in Figure 1.

If $\Delta\phi(x,y,z)$ is the head-detected flux arising from a magnetic source of the particle moment component, m, at point (x,y,z) in the tape, and if ΔN is the noise flux arising from a volume of the tape, Δv, which is large enough to contain a statistical sample of the particles but small enough so that $\phi(x,y,x)$ varies negligibly over it (~ 0.5 mil^3), then

$$(\Delta N)^2 = \int_{\Delta v} (\Delta\phi - \bar{\Delta\phi})^2 \, dn_o$$

$$= n \Delta v \int_{\Delta v} (\Delta\phi - \bar{\Delta\phi})^2 \frac{dn_o}{n \Delta v} \quad \ldots\ldots (1a)$$

where n is the average particle density within Δv. Although n', the particle density at a point, is a random variable and thus varies within Δv, in a homogeneous tape, n itself is a constant for all Δv.

$$(\Delta N)^2 = n \Delta v \, E\left[(\Delta\phi - \bar{\Delta\phi})^2\right] \quad \ldots\ldots (1b)$$

where $E\left[(\Delta\phi - \bar{\Delta\phi})^2\right]$ is the expectation of $(\Delta\phi - \bar{\Delta\phi})$.

Then, since by definition the expectation of $(\Delta\phi - \bar{\Delta\phi})$ is the variance of $\Delta\phi \left(\sigma^2(\Delta\phi)\right)$, we have

$$(\Delta N)^2 = n \Delta v \, \sigma^2(\Delta\phi) \quad \ldots\ldots (1c)$$

(Note: $\frac{dn_o}{n \Delta v}$ is the probability distribution of $\Delta\phi$.)

It is interesting to note that the randomness of the particle density in a homogeneous tape does not contribute to the noise, except as a second order effect. Thus, with respect to noise, a perfectly uniform array, random only in the particle direction, is equivalent to the given randomly dispersed array. The arrangement of the array becomes significant only if the moments and their directions are all identical, in which case the first order effects disappear. The effect of random arrangement can be calculated by assuming that each statistical ensemble consists of identical numbers of particles, of which a small random number are of zero moment or oriented in a direction such that their fields are not detected by the reproduce head. Thus, the effect of a random distribution is merely to modify slightly the distribution function of the particles. Leaving aside this second-order effect, the noise from the entire tape is

$$N^2 = \int (\Delta N)^2 = n \int \sigma^2(\Delta\phi) \Delta v \quad \ldots\ldots (1d)$$

It is thus necessary to evaluate $\sigma^2(\Delta\phi)$.

The flux in a reproduce head due to a magnetization $\vec{M}$ in a tape is given by the formula[2]:

$$\phi = \int_v \vec{M} \cdot \vec{H} \, dv \quad \ldots\ldots (2a)$$

where $\vec{H}$ is the field from an identical head having a unit difference of potential on its two surfaces.

This formula holds as long as the permeability in the head is independent of field, i.e. as long as the field relationships are linear. This is certainly true for the essentially infinite permeability heads normally used. It should be noted that this formula does not consider the effect of the reproduce head on the orientation of the tape particles, an effect which we do not evaluate in this investigation. Then, the head-detected flux from a single particle of the tape is:

$$\Delta\phi(x,y,z) = \vec{M} \cdot \vec{H} \, \Delta v_o \quad \ldots\ldots (2b)$$

where Δv_o is the average volume per particle in the particle-matrix mix, and $\vec{M}$ is the magnetization in Δv.

Then, if each particle has a volume u, a magnetization $\vec{j}$, and a magnetic moment $\vec{m}$, we have:

$$\vec{M} \, \Delta v_o = \vec{j} \left(\frac{u}{\Delta v_o}\right) \Delta v_o = \vec{j} \, u = \vec{m} \quad \ldots\ldots (3)$$

and

$$\Delta\phi(x,y,z) = \vec{m} \cdot \vec{H} = m H \cos(m,H) \quad \ldots\ldots (2c)$$

Assuming that H_x and H_z do not affect m_z and m_x, respectively (which is only approximately true), then[3]:

$$\sigma^2(\Delta\phi) = \sigma^2(\vec{m} \cdot \vec{H}) = \sigma^2(m_x H_x + m_z H_z)$$

$$= H_x^2 \, \sigma^2(m_x) + H_z^2 \, \sigma^2(m_z)^2 \quad \ldots\ldots (4)$$

Then, from Equations (1) and (4) we have:

$$N^2 = n \int_v \left[H_x^2 \, \sigma^2(m_x) + H_z^2 \, \sigma^2(m_z)\right] dv$$

or

$$N^2 = n \left[\sigma^2(m_x) \int_v H_x^2 \, dv + \sigma^2(m_z) \int_v H_z^2 \, dv\right] \quad \ldots\ldots (5)$$

Now we have:

$$m_x = m \cos(m,x) \equiv m \beta_x$$

$$m_z = m \cos(m,z) \equiv m \rho_z \quad \ldots\ldots (6)$$

Then:

$$\sigma^2(m_x) = m^2\ \sigma^2(\beta_x) = m^2\ \overline{\beta_x}^2$$

$$\sigma^2(m_z) = m^2\ \sigma^2(\beta_z) = m^2\ \overline{\beta_z}^2 \quad \ldots\ldots (7)$$

Therefore:

$$N^2 = nm^2 \left[\overline{\beta_x}^2 \int_v H_x^2\, dv + \overline{\beta_z}^2 \int_v H_z^2\, dv \right] \ldots (5a)$$

Now, in the section on spectrum analysis of noise, we show that

$$\int_v H_x^2\, dv = \int_v H_z^2\, dv \quad \ldots\ldots\ldots (8)$$

Therefore:

$$N^2 = \frac{nm^2}{2} \left(\overline{\beta_x}^2 + \overline{\beta_z}^2 \right) \int_v H^2\, dv \quad \ldots\ldots (9a)$$

or:

$$N^2 = \frac{nm^2}{2} \left(\overline{\beta}^2 \right) \int_v H^2\, dv \quad \ldots\ldots\ldots (9b)$$

Now, since the reproduce head structure is assumed to be uniform across the track width and of large dimensions compared with the thickness and the statistical region, Δu, then H_x and H_z are not functions of the lateral coordinate y.
Therefore:

$$\int_v H^2\, dv = w \int_{-\infty}^{\infty} dx \int_d^{d+t} H^2(x,z)\, dz \quad \ldots\ldots (10a)$$

If $\overline{H^2}$ is the mean square of H, then:

$$\int_v H^2\, dv = wt \int_{-\infty}^{\infty} \overline{H^2}(x,t,d)\, dx \quad \ldots\ldots (10b)$$

Since it is generally true that the second moment, N^2, is proportional to the total number in the collection, we now wish to put our formula for N^2 in such a form that it is porportional to the 'total number' of particles the head detects. This requires the introduction of effective lengths of tape for each field component. Therefore, if x_1 is the location of maximum H_x, and x_2 is the location of maximum H_z, then:

$$\overline{G_x}^2(t,d) \equiv \int_{-\infty}^{\infty} \overline{H_x}^2(x,t,d)\, dx = r_x \overline{H_x}^2(x_1,t,d)$$

$$\overline{G_z}^2(t,d) \equiv \int_{-\infty}^{\infty} \overline{H_z}^2(x,t,d)\, dx = r_z \overline{H_z}^2(x_2,t,d) \quad \ldots (11)$$

where r_x and r_z (it will be shown) are measures of the effective lengths of tape seen by the head in noise detection of the x and z components of the particle moments.

Then:

$$N^2 = n\ wt\ m^2 \left[\overline{\beta_x}^2 r_x \overline{H_x}^2(x_1,t,d) + \overline{\beta_z}^2 r_z \overline{H_z}^2(x_2,t,d) \right] \quad \ldots\ldots\ldots (12)$$

In order to interpret r_x and r_z, we refer to Equation (2a), where the flux, $\Delta\phi$, can also be written as:

$$\Delta\phi(x,y,z) = m_o g(x,y,z) \quad \ldots\ldots\ldots (13)$$

where $g(x,y,z)$ is the flux detected by the head from a unit source at a point (x,y,z) in the tape, and m_o is the strength of the source at (x,y,z). Thus, m_o can be identified with m cos (m H), and $g(x,y,z)$ with $H(x,y,z)$. Therefore, H_x and H_z are the x and z components of the flux from a unit source in the tape detected by the head; and r_x and r_z are measures of effective lengths of tape seen by the head in noise (mean square flux) detection of the x and z components of the particle moments.

We can regroup the terms in Equation (12) as:

$$N^2 = N_1^2 + N_2^2 = \left[m^2 (n\ wt\ r_x)\ \overline{\beta_x}^2\ \overline{H_x}^2(x_1,t,d) \right] + \left[m^2 (n\ wt\ r_z)\ \overline{\beta_z}^2\ \overline{H_z}^2(x_2,t,d) \right] \quad \ldots\ldots (14)$$

Thus, the noise detected by the head has two independent sources, the x and z components of the particle moments. It should be pointed out that the definition of r_x and r_z in terms of $\overline{H_x}^2$ maximum and $\overline{H_z}^2$ maximum is somewhat arbitrary. The definition could just as well have been in terms of $\overline{H_x}^2$ average and $\overline{H_z}^2$ average, or $\overline{H_x}^2 + \overline{H_z}^2$ maximum.

The expression for the noise in Equation (14) is quite general. It states that a detected noise power, say N_1^2 in a magnetic head, from a random set of sources, e.g., in a magnetic tape, is proportional to the sum of noise sources, n wt r_x, each of strength $\left[m^2\ \overline{\beta_x}^2\ \overline{H_x}^2(x,t,d) \right]$. The noise source itself can be further analyzed. It is the product of three terms: (1) $\overline{\beta_x}^2$, the value of the noise source; (2) m^2, the strength of the noise source; and (3) $\overline{H_x}^2(x_1,t,d)$, the mean square flux from an 'average' unit magnetic source. (Of course, the same statement also holds for N_2^2.)

Now, $r_x, r_z = f(t,d)$; thus, the number of noise sources the head 'sees' or detects from each component is a function of both the tape thickness and the head-tape spacing. Since $\overline{H_x}^2(x_1,t,d)$ and $\overline{H_z}^2(x_2,t,d)$ (both mean square flux from an 'average' unit magnetic source)

are also a function of t and d, then the noise source strength, m^2, is a function of t and d. Thus, the head-detected noise will vary with tape thickness and tape-head spacing because, generally, both the number of effective noise sources and their detected strengths vary. As the head-tape spacing increases, the noise, of course, will drop off, even though we expect the number of particles the head 'sees' to increase. Thus, the noise and the noise resolutions, r_x and r_z, have an inverse relationship to each other, as will be shown later on. Furthermore, since the head-detected flux is a function of the head structure, then $G^2(t,d)$ and therefore N^2 are functions of the head structure. This statement will be illustrated by specific examples when we evaluate the noise from specific head-tape systems.

It is now of considerable interest to compare the no-signal noise, which is the limit of the 'lowest possible' recorded level, with the output of a saturated tape, which is a measure of the maximum possible recorded level. The d.c. signal, or remanent flux from a saturated tape, is simply the flux created in a cross-section of the tape. However, for purposes of comparison with the noise formula, we will derive the simple formula for S in a manner analogous with the derivation of N^2. Thus, we have:

$$S = \int (\Delta\phi)\, dn = n \int (\Delta\phi)\, dv \quad \ldots\ldots (15a)$$

$$= n\, m\, wt \left[\overline{\cos}\,(m,x) \int_{-\infty}^{\infty} \bar{H}_x(t,d,x)\, dx + \overline{\cos}(m,z) \int_{-\infty}^{\infty} \bar{H}_z(t,d,x)\, dz \right] \quad \ldots\ldots\ldots (15b)$$

Or, in other notation:

$$S = n\, m\, wt \left[\bar{\beta}_x \bar{G}_x(t,d) + \bar{\beta}_z \bar{G}_z(t,d) \right] \quad \ldots (15c)$$

where

$$\bar{G}_x(t,d) = \int_{-\infty}^{\infty} \bar{H}_x(t,d,x)\, dx = \int_{-\infty}^{\infty} \frac{dx}{t} \int_{d}^{d+t} H_x(x,z)\, dz$$

$$\bar{G}_z(t,d) = \int_{-\infty}^{\infty} \bar{H}_z(t,d,x)\, dx = \int_{-\infty}^{\infty} \frac{dx}{t} \int_{d}^{d+t} H_z(x,z)\, dz \quad \ldots\ldots (16)$$

Since $\bar{G}_z(t,d) = 0$ for a symmetrical head, we have:

$$S = n\, m\, wt\ \bar{\beta}_x \bar{G}_x \quad \ldots\ldots\ldots\ldots (15d)$$

Thus, the flux entering the head is simply the flux created in a cross-section of the tape.

We can define resolution terms for the d.c. signal the same as for the noise. Thus:

$$R_x \bar{H}_x(x,t,d) = \bar{G}_x \quad \ldots\ldots\ldots\ldots (17)$$

and

$$S = (n\, wt\, R_x)\, m\ \bar{\beta}_x \bar{H}_x(x,t,d) \quad \ldots\ldots\ldots (18)$$

This is quite general and is simply the formula for the coherent addition of sources.

Comparing S and N, we get:

$$\frac{S^2}{N^2} = 2\, n\, wt \left[\frac{\left(\bar{\beta}_x \bar{G}_x\right)^2}{\overline{\beta^2}\ \overline{G_x^2}} \right] \quad \ldots\ldots\ldots (19a)$$

or

$$\frac{S^2}{N^2} = 2n\, wt \left[\frac{\left(\bar{\beta}_x R_x \bar{H}_x(x_1,t,d) \right)^2}{\beta^2 \left(r_x \overline{H_x^2}(x_1,t,d) + r_z \overline{H_z^2}(x_2,t,d) \right)} \right] \quad (19b)$$

and

$$\frac{S^2}{N^2} = \left(n\, wt\, \frac{R^2}{r} \right) \left[\frac{\left(\bar{\beta}_x \cos\theta_1 \bar{H}_x(x_1,t,d) \right)^2}{\overline{\beta_x^2} \cos\theta_2 \overline{H_x^2}(x_1,t,d) + \overline{\beta_z^2} \sin\theta_2 \overline{H_z^2}(x_2,t,d)} \right] \quad \ldots\ldots (19c)$$

where $r^2 \equiv r_x^2 + r_z^2$; $\theta_2 \equiv \tan^{-1}\left(\frac{r_z}{r_x}\right)$, and similarly for R and θ_1.

We see that the ratio of signal to noise powers is proportional to an effective number of source particles $\left(n\, wt\, \frac{R^2}{r}\right)$ and the ratio of average to second-moment terms. This general form, of course, is not contingent upon the particular physical system we are concerned with, but is the result of calculating the ratio of the first to the second moment of some property of any collection of random independent sources. However, by putting S/N into this general form, we can more clearly envision the physical basis of S/N statistical characteristics for tape-head systems. Thus, the concepts of noise and signal resolutions, i.e. r and R, become apparent if we express S/N in a general statistical formulation like the above.

Interestingly enough, the resolutions themselves can be expressed as statistical moments of the unit field components, H_x and H_z. This appears reasonable,

since certainly the effective length of tape measured in noise and dc signal determinations will be a function of the head field. And since the head does not detect the flux from the particles individually but only as a sum, any spatial ordering of the fluxes from the particles is irrelevant to the head. Thus we write:

$$\frac{\bar{G}_x}{\overline{G_x^2}} = \lim_{a\to\infty} \frac{\int_{-\infty}^{\infty} \bar{H}_x(x,t,d)\,\frac{dx}{a}}{\int_{-\infty}^{\infty} \overline{H_x^2}(x,t,d)\,\frac{dx}{a}} = \frac{E(\bar{H}_x)}{E\left(\overline{H_x^2}\right)} \quad \ldots\ldots (20)$$

where $\frac{1}{a}$ is the probability function of a particle in a homogeneous tape of a length dx. It is to be noted that both $E(H_x)$ and $E(H_x^2)$ are zero, but the ratio is defined and has the value $\frac{\bar{G}_x}{\overline{G_x^2}}$. Then:

$$\frac{E(\bar{H}_x)}{E\left(\overline{H_x^2}\right)} = \frac{(\bar{H}_x)_{av}}{\sigma^2(\bar{H}_x) + \left[(\bar{H}_x)_{av}\right]^2}$$

$$= \frac{1}{\frac{\sigma^2(\bar{H}_x)}{(\bar{H}_x)_{av}} + (\bar{H}_x)_{av}} \quad \ldots\ldots (21a)$$

Therefore, since $(\bar{H}_x)_{av} = 0$, then:

$$\frac{E(\bar{H}_x)}{E\left(\overline{H_x^2}\right)} = \frac{(\bar{H}_x)_{av}}{\sigma^2(\bar{H}_x)} \quad \ldots\ldots\ldots (21b)$$

Thus,

$$\frac{r_x}{R_x} = \frac{\frac{\sigma^2(\bar{H}_x)}{\overline{H_x^2}(x_1)}}{\frac{(\bar{H}_x)_{av}}{\bar{H}_x(x_1)}}$$

and

$$\frac{r_z}{R_x} = \frac{\frac{\sigma^2(\bar{H}_z)}{\overline{H_z^2}(x_1)}}{\frac{(\bar{H}_x)_{av}}{\bar{H}_x(x_1)}} \quad \ldots\ldots\ldots (22)$$

Thus, the ratio of the resolutions can be written as a 'normalized noise source'. We can also express S/N in terms of this 'noise source':

$$\frac{S^2}{N^2} = n\,wt \left[\frac{(\bar{\beta}_x)^2}{\overline{\beta_x^2}\left(\frac{\sigma^2(\bar{H}_x)}{(\bar{H}_x)_{av}}\right) + \overline{\beta_z^2}\left(\frac{\sigma^2(\bar{H}_z)}{(\bar{H}_z)_{av}}\right)} \right] \quad (23)$$

4. Evaluation of the Field Terms

It is now necessary to actually calculate S/N. We will first evaluate the field terms and thus show explicitly how S/N depends upon the head-tape system. As our first system we chose a simple but useful case, i.e. a zero magnetic gap or any zero thickness linear detector of infinite extent. The field components arising from a unit difference potential head (i.e., the head-detected flux from unit source components) are[2]:

$$H_x = \frac{4z}{x^2+z^2}$$

$$H_z = \frac{4x}{x^2+z^2} \quad \ldots\ldots\ldots\ldots (24)$$

Then

$$\bar{G}_x = \int_{-\infty}^{\infty} \bar{H}_x(x,t,d)\,dx = \int_{-\infty}^{\infty} \frac{dx}{t} \int_d^{d+t} H_x(x,z)\,dz$$

$$= \frac{4}{t}\int_d^{d+t} z\,dz \int_{-\infty}^{\infty} \frac{dx}{x^2+z^2} = 4\pi$$

$$\overline{G_x^2} = \int_{-\infty}^{\infty} \overline{H_x^2}(x,t,d)\,dx = \int_{-\infty}^{\infty} \frac{dx}{t} \int_d^{d+t} H_z^2(x,z)\,dz$$

$$= \frac{16}{t}\int_d^{d+t} z^2\,dz \int_{-\infty}^{\infty} \frac{dx}{(x^2+z^2)^2} = \frac{8\pi}{t}\log\left(1+\frac{t}{d}\right) \quad \ldots\ldots (25a)$$

$$\overline{G_z^2} = \int_{-\infty}^{\infty} \overline{H_z^2}(x,t,d)\,dx = \int_{-\infty}^{\infty} \frac{dx}{t} \int_d^{d+t} H_z^2(x,z)\,dz$$

$$= \frac{16}{t}\int_d^{d+t} dz \int_{-\infty}^{\infty} \frac{x^2\,dx}{(x^2+z^2)^2} = \frac{8\pi}{t}\log\left(1+\frac{t}{d}\right) \quad \ldots\ldots (25b)$$

Thus, from Equations (5a) and (25) we get:

$$N^2 = n\,wt\,m^2\left[\overline{\beta_x^2} + \overline{\beta_z^2}\right]\frac{16\pi}{t}\log\left(1+\frac{t}{d}\right) \quad \ldots (26)$$

and from Equations (15d) and (25) we get:

$$S = n\,wt\,m\left(\bar{\beta}_x\right) 4\pi \quad \ldots\ldots\ldots (27)$$

Therefore,

$$\frac{S^2}{N^2} = n\,wt\left[\frac{\pi t}{\lg\left(1+\frac{t}{d}\right)}\right]\left[\frac{\left(\bar{\beta}_x^2\right)}{\bar{\beta}_x^2 + \bar{\beta}_z^2}\right] \quad \ldots\ldots (28)$$

For $\frac{t}{d} << 1$,

$$N^2 \cong n\,wt\,m^2\left(\bar{\beta}^2\right)\frac{16\pi}{d} \quad \ldots\ldots\ldots (29)$$

$$\frac{S^2}{N^2} \cong n\,wt\left[\frac{\left(\bar{\beta}_x\right)^2}{\bar{\beta}^2}\right]\pi d \quad \ldots\ldots (30)$$

Thus, for thin tapes spaced far away from the head, $\frac{S^2}{N^2}$ is proportional to both thickness and spacing, while N^2 is proportional to $\frac{t}{d}$. For larger values of $\frac{t}{d}$, $\frac{S^2}{N^2}$ increases much more rapidly with thickness but only slightly more rapidly with spacing, while N^2 is affected only slightly by variations in either t or d. Generally, the variation of $\frac{S^2}{N^2}$ with thickness and spacing can be interpreted as the effective number of particles the head 'sees'.

The formulas for N^2 and $\frac{S^2}{N^2}$ are plotted in Figure 2. We now calculate the resolutions. From Equation (24) we have:

$$\bar{H}_x^2(x_1,t,d) = \frac{1}{t}\int_d^{d+t} H_x^2(x_1,z)\,dz$$

$$= \frac{16}{t}\int_d^{d+t}\frac{dz}{z^2} = \frac{16}{d(d+t)},\ x_1 = 0 \quad \ldots (31a)$$

$$\bar{H}_z^2(x_2,t,d) = \frac{1}{t}\int_d^{d+t} H_z^2(x_2,z)\,dz$$

$$= \frac{4}{t}\int_d^{d+t}\frac{dz}{z^2} = \frac{4}{d(d+t)},\ x_2 = z \quad \ldots (31b)$$

$$\bar{H}_x(x_1,t,d) = \frac{1}{t}\int_d^{d+t} H_x(x_1,z)\,dz$$

$$= \frac{4}{t}\int_d^{d+t}\frac{dz}{z} = \frac{4}{t}\lg\left(1+\frac{t}{d}\right) \quad \ldots\ldots (31c)$$

$$\bar{H}_z(x_2,t,d) = \frac{1}{t}\int_d^{d+t} H_z(x_2,z)\,dz$$

$$= \frac{2}{t}\int_d^{d+t}\frac{dz}{z} = \frac{2}{t}\lg\left(1+\frac{t}{d}\right) \quad \ldots\ldots (31d)$$

Therefore, from Equation (11), the resolutions are:

$$r_x = \frac{\bar{G}_x^2}{\bar{H}_x^2(x_1)} = \left(\frac{1}{2}\right)\frac{\pi d\left(1+\frac{t}{d}\right)}{\frac{t}{d}}\lg\left(1+\frac{t}{d}\right) \quad \ldots (32a)$$

$$r_z = \frac{\bar{G}_z^2}{\bar{H}_z^2(x_1)} = \frac{2\pi d\left(1+\frac{t}{d}\right)}{\frac{t}{d}}\lg\left(1+\frac{t}{d}\right) \quad \ldots (32b)$$

$$R_x = \frac{\bar{G}_x}{\bar{H}_x(x_1)} = \frac{\pi t}{\lg\left(1+\frac{t}{d}\right)} \quad \ldots\ldots\ldots (32c)$$

The formulas for the resolutions are plotted in Figure 3. We notice that $r_z = 4\,r_x$, indicating that the noise length from the z component of the particles, arising from an effective length of tape detected by the head, is four times as great as the noise length from the x component. This is not surprising, since the field from the x component is more closely confined to the gap region than the field from the z component.

It is easily seen that $R_x \geq r_x$ for all t and d, and that $R_x \geq r_z$ for all t and d except $\frac{t}{d} << 1$. The effect of an incoherent, as compared to a coherent, superposition of signals, therefore, is to decrease the effective length of the tape detected by the head.

For $\frac{t}{d} << 1$,

$$r_x \cong \frac{\pi d}{2}$$

$$r_z \cong 2\pi d$$

$$R_x \cong \pi d \quad \ldots\ldots\ldots\ldots (33)$$

For $\frac{t}{d} >> 1$,

$$r_x \cong \frac{\pi d}{2} \lg\left(\frac{t}{d}\right)$$

$$r_z \cong 2\pi d \lg\left(\frac{t}{d}\right)$$

$$R_x \cong \frac{\pi t}{\lg\left(\frac{t}{d}\right)} \quad \ldots\ldots\ldots\ldots (34)$$

Calculating the resolution for the actual field around a non-zero gap head is a tedious problem. However, since the field shape is known, the effect of the actual field on $\frac{S^2}{N^2}$ can be estimated. Since the field is approximately uniform in the region $\frac{\ell}{8} \leq z \leq \ell$, S/N approaches ℓ there. In the region $0 \leq z \leq \frac{\ell}{8}$, the field is peaked at the edges. Therefore, as $z \longrightarrow 0$, all resolutions approach zero; but $\bar{H}_x(x_1)$ and $\bar{H}_z(x_2)$ grow larger, and therefore N^2 grows larger. (Theoretically, $N^2 \longrightarrow \infty$.) Thus, even though $N^2 \longrightarrow \infty$, the smaller the spacing, the smaller the noise resolution and the larger the S/N.

It is interesting to note that for a zero-gap infinite head, H(y,z) is also the field from a row of unit particles across the width of the tape. This is to be expected, since within an infinite permeability head the lines of magnetic induction from the tape field are the same as with no head present.

5. Noise Spectrum Analysis

Now, a complete specification of the random variations in a system, which we refer to as noise, requires the knowledge of an infinite set of terms, such as the values of all the statistical moments or a complete spectral analysis. The definition of noise as the second moment of the statisitical variable (the noise power), therefore, is a significant but not complete statement of the random variations. The most useful expression of completeness is the specification of the spectral density.

The spectrum of a random assemblage of independent identical sources, of course, is the same as that of the individual sources. However, since the sources in the tape are spaced varying amounts from the head, the field (and thus the spectral density of the field) will vary for particles throughout the tape thickness. We thus expect the spectral density of the head-detected N to be a function generally of both tape thickness and head-tape spacing. Furthermore, since the field from the sources is uniquely determined by the field boundary conditions established by the head geometry, we expect the spectral density of N to be a function of head geometry. Thus, the noise at a given wavelength is detected similarly as a signal of the same wavelength, the head, of course, being incapable of distinguishing between noise and signal. The S/N at a given wavelength, therefore, is not a function of the head-tape spacing or of the head geometry. A general proof of this statement will be given later in this section. We now determine the noise and signal resolutions at a given wavelength and also the narrow band signal-to-noise.

Equation (5a) states:

$$N^2 = n\ wt\ m^2 \left[\overline{\beta_x^2}\ \overline{G_x^2}(t,d) + \overline{\beta_z^2}\ \overline{G_z^2}(t,d) \right]$$

where from Equation (11),

$$\overline{G_x^2}(t,d) = \frac{1}{t}\int_d^{d+t} dz \int_{-\infty}^{\infty} H_x^2(x,z)\, dx$$

$$\overline{G_z^2}(t,d) = \frac{1}{t}\int_d^{d+t} dz \int_{-\infty}^{\infty} H_z^2(x,z)\, dx$$

Now, since it is assumed there is no particle interaction, then:

$$\int_0^{\infty} H_x^2(x,z)\, dx = \int_0^{\infty} C_x^2(k,z)\, dk$$

$$\int_0^{\infty} H_z^2(x,z)\, dx = \int_0^{\infty} C_z^2(k,z)\, dk \quad \ldots\ldots (35)$$

where k is the wavenumber ($\equiv \frac{2\pi}{\lambda}$) and C_x and C_z, the spectral densities, are:

$$C_x(k,z) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} H_x(x,z) \cos kx\, dx$$

$$C_z(k,z) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} H_z(x,z) \sin kx\, dx \quad \ldots\ldots (36)$$

Then, the noise power in bandwidth dk is:

$$d(N^2) = 2n\ wt\ m^2 \left[\overline{\beta_x^2}\ \overline{C_x^2}(k,t,d) + \overline{\beta_z^2}\ \overline{C_z^2}(k,t,d) \right] dk \quad \ldots (37)$$

where

$$\overline{C_x^2}(k,t,d) = \frac{1}{t}\int_d^{d+t} C_x^2(k,z)\, dz$$

$$\overline{C_z^2}(k,t,d) = \frac{1}{t}\int_d^{d+t} C_z^2(k,z)\, dz \quad \ldots (38)$$

Since it will be shown later that $C_x = C_z$, then from Equations (38) and (11), the spectral density of the resolution is given by:

$$dr_x = \left[\frac{2\overline{C_x^2}(k,t,d)}{\overline{H_x^2}(x_1)}\right] dk$$

$$dr_z = \left[\frac{2\overline{C_x^2}(k,t,d)}{\overline{H_z^2}(x_2)}\right] dk \quad \ldots\ldots\ldots (39)$$

We now evaluate R_x and R_z, the resolutions of a given recorded wavelength. From Equation (17) we have:

$$R_x = \frac{\bar{G}_x}{\bar{H}_x(x_1,t,d)} = \frac{\int_{-\infty}^{\infty} \bar{H}_x(t,d,x) \cos kx\, dx}{\bar{H}_x(x_1,t,d)}$$

$$R_z = \frac{\bar{G}_z}{\bar{H}_z(x_1,t,d)} = \frac{\int_{-\infty}^{\infty} \bar{H}_z(t,d,x) \sin kx\, dx}{\bar{H}_z(x_2,t,d)} \quad \ldots (40)$$

Since $C_x = C_z$, $G_x = \sqrt{2\pi}\, C_x$, and $G_z = \sqrt{2\pi} C_z$, then we have, from Equation (15c), the peak signal:

$$S = n\ m\ wt \left[(\bar{\beta}_x)^2 + (\bar{\beta}_z)^2 \right]^{1/2} \bar{G}_x$$

$$= \sqrt{2\pi}\ n\ m\ wt \left[(\bar{\beta}_x)^2 + (\bar{\beta}_z)^2 \right] \bar{C}_x \quad (41)$$

Thus:

$$\frac{S^2}{\frac{d(N^2)}{dk}} = \pi n\ wt\ \frac{\left[(\bar{\beta}_x)^2 + (\bar{\beta}_z)^2 \right]}{\overline{\beta^2}}\ \frac{(\bar{C}_x)^2}{\overline{C_x^2}} \quad (42a)$$

In terms of the resolution spectrum:

$$\frac{S^2}{\frac{d(N^2)}{dk}} = 2\pi n\ wt\ \frac{\left[(\bar{\beta}_x) + (\bar{\beta}_z)^2 \right]}{\overline{\beta^2}}$$

$$\frac{\left[R_x \bar{H}_x(x_1) + \bar{R}_z \bar{H}_z(x_2) \right]^2}{\overline{H_x^2}(x_1)\, dr_x + \overline{H_z^2}(x^2)\, dr_z} \quad \ldots\ldots (42b)$$

We now determine the spectrum using the field equation for the zero-gap head. We have:

$$C_x(k,z) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} \frac{4z}{x^2 + z^2} \cos kx\, dx$$

$$= 4\sqrt{\frac{\pi}{2}}\ \ell^{-kz}$$

$$C_z(k,z) = \sqrt{\frac{2}{\pi}} \int_0^{\infty} \frac{4x}{x^2 + z^2} \sin kx\, dx$$

$$= 4\sqrt{\frac{\pi}{2}}\ \ell^{-kz} \quad \ldots\ldots (43)$$

(If the head is finite, then for small k, $C_z(k,z) \longrightarrow 0$.)

Therefore:

$$\frac{1}{t} \int_d^{d+t} C_x^2(k,z)\, dz = \frac{8\pi}{t} \int_d^{d+t} e^{-2kz}\, dz$$

$$= \frac{8\pi e^{-2kd}\left[1 - e^{-2kt}\right]}{2kt}$$

and also:

$$\frac{1}{t} \int_d^{d+t} C_z^2(k,z)\, dz = \frac{8\pi e^{-2kd}\left[1 - e^{-2kt}\right]}{2kt}$$

$$\ldots\ldots\ldots (44)$$

Therefore:

$$d(N^2) = n\ wt\ m^2\ \beta^2 \left[\frac{16\pi e^{-2kd}\left[1 - e^{-2kt}\right]}{2kt}\right] dk$$

$$\ldots\ldots\ldots (45)$$

and since

$$\frac{1}{t} \int_d^{d+t} C_x(k,z)\, dz = \frac{4}{t}\sqrt{\frac{\pi}{2}} \int_d^{d+t} e^{-kz}\, dz$$

$$= 4\sqrt{\frac{\pi}{2}} \left[\frac{e^{-kd}\left[1 - e^{-kt}\right]}{kt}\right]$$

$$\frac{1}{t} \int_d^{d+t} C_z(k,z)\, dz = \frac{4}{t}\sqrt{\frac{\pi}{2}} \left[\frac{e^{-kd}\left[1 - e^{-kt}\right]}{kt}\right] \quad (46)$$

we have for the root mean square signal:

$$S = 4\pi n\ wt\ m \left[\frac{(\bar{\beta}_x)^2 + (\bar{\beta}_z)^2}{\sqrt{2}}\right]^{1/2} \left[\frac{e^{-kd}\left[1 - e^{-kt}\right]}{kt}\right] \quad (47)$$

and

$$\frac{S^2}{\frac{d(N^2)}{dk}} = \frac{\pi}{2}\, n\, wt\, \frac{\left[(\bar{\beta}_x)^2 + (\bar{\beta}_z)^2\right]}{\overline{\beta^2}} \left(\frac{\tanh\left(\frac{kt}{2}\right)}{\frac{kt}{2}}\right)^2 \quad \ldots\ldots (48a)$$

For uniaxial particles randomly oriented, it will be shown that

$$\bar{\beta}_x = \bar{\beta}_z = \frac{1}{2\sqrt{2}}, \text{ and } \overline{\beta^2} = \frac{2}{3}.$$ Therefore[4,5]:

$$\frac{S^2}{\frac{d(N^2)}{dk}} = \frac{3\pi}{16}\, n\, wt \left[\frac{\tanh\left(\frac{kt}{2}\right)}{\frac{kt}{2}}\right] \quad \ldots\ldots\ldots (48b)$$

Equation 48a is plotted in Figure 4.

For long wavelengths or small tape thicknesses:

$$\frac{S^2}{\frac{d(N^2)}{dk}} \cong \frac{\pi}{2}\, n\, wt\, \frac{\left[(\bar{\beta}_x)^2 + (\bar{\beta}_z)^2\right]}{\overline{\beta^2}} = \frac{3\pi}{16}\, n\, wt \quad \ldots\ldots (49a)$$

Thus, in this region, the narrow band S^2/N^2 is proportional to the tape cross-sectional area.

For short wavelengths or large tape thicknesses:

$$\frac{S^2}{\frac{d(N^2)}{dk}} \cong \frac{\pi}{2}\, n\, wt\, \frac{\left[(\bar{\beta}_x)^2 + (\bar{\beta}_z)^2\right]}{\overline{\beta^2}} \left(\frac{1}{\frac{kt}{2}}\right) = \frac{3\pi}{16}\, n\, wt \left(\frac{1}{\frac{kt}{2}}\right) \quad \ldots\ldots\ldots (49b)$$

and the narrow band S^2/N^2 is independent of the thickness and proportional to the wavelength. A tentative comparison of theory with unevaluated experiments shows close correspondence. It will be shown in the section on d.c. noise that d.c. noise can be understood if we assume that only a small fraction of the particles contribute to the no-signal noise, and that therefore there must be sources of noise in the tape not evaluated here.

The spectral analysis of the resolutions gives:

$$dr_x = \left[\frac{\overline{C_x^2}(k,t,d)}{\overline{H_x^2}(x_1)}\right] dk = \left[\frac{2\pi e^{-2kd}\left(1-e^{-2kt}\right) d(d+t)}{2kt}\right] dk$$

$$dr_z = \left[\frac{\overline{C_z^2}(k,t,d)}{\overline{H_z^2}(x_2)}\right] dk = \left[\frac{8\pi e^{-2kd}\left(1-e^{-2kt}\right) d(d+t)}{2kt}\right] dk$$

$$R_x = \frac{\bar{C}_x(k,t,d)}{\bar{H}_x(x_1)} = \sqrt{\frac{\pi}{2}} \left[\frac{e^{-kd}\left[1-e^{-kt}\right] t}{kt \lg\left(1+\frac{t}{d}\right)}\right]$$

$$R_z = \frac{\bar{C}_z(k,t,d)}{\bar{H}_z(x_2)} = \frac{1}{2}\sqrt{\frac{\pi}{2}} \left[\frac{e^{-kd}\left[1-e^{-kt}\right] t}{kt \lg\left(1+\frac{t}{d}\right)}\right] \quad \ldots (50)$$

Equation (48) shows that for the zero-gap field, the narrow band-signal-to-noise is independent of head-to-tape spacing. Thus, the spectral density, or the 'number' of wavenumbers the head 'sees', is independent of the spacing. This is obvious, since the range of any wavelength is $(-\infty, \infty)$. Therefore, we expect the narrow band S/N to be independent of gap width, at least for an infinite head of effectively infinite permeability. This can now be shown explicitly.

The potential in the region outside an infinite head of effectively infinite permeability (i.e., where $z \geq 0$) can be expressed in terms of a Fourier integral[6]:

$$\phi(x,z) = \sqrt{\frac{2}{\pi}} \int_0^\infty g(k) \sin kx\, e^{-kz}\, dz \quad \ldots (51)$$

This, of course, is nothing but the required solution of the Laplace equation, which describes the field in the head region. Then:

$$H_x(x,z) = -\frac{\partial \phi}{\partial x} = \sqrt{\frac{2}{\pi}} \int_0^\infty kg(k) \cos kx\, e^{-kz}\, dk$$

$$H_z(x,z) = -\frac{\partial \phi}{\partial z} = \sqrt{\frac{2}{\pi}} \int_0^\infty kg(k) \sin kz\, e^{-kz}\, dk \quad (52a)$$

Therefore,

$$kg(k)\, e^{-kz} = \sqrt{\frac{2}{\pi}} \int_0^\infty H_x(x,z) \cos kx\, dx$$

and

$$kg(k)\, e^{-kz} = \sqrt{\frac{2}{\pi}} \int_0^\infty H_z(x,z) \sin kz\, dx \quad \ldots (52b)$$

However, by Equation (36) we see that

$$\left|C_x(k,z)\right| = \left|C_z(k,z)\right| = k\, g(k)\, e^{-kz}$$

Therefore:

$$\overline{C^2} = \overline{C_x^2} + \overline{C_z^2} = k^2 g^2(k) \frac{2}{t} \int_d^{d+t} e^{-2kz}\, dz$$

$$= 2k^2 g^2(k) \frac{e^{-2kd}\left[1-e^{-2kt}\right]}{2kt} \quad \ldots\ldots\ldots (53a)$$

also:

$$\bar{C}^2 = \bar{C}_x^{\,2} + \bar{C}_z^{\,2} = 2k^2g^2(k)\left[\frac{e^{-2kd}\left[1-e^{-kt}\right]2}{k^2t^2}\right] \quad . \quad (53b)$$

Thus, from Equation (42a),

$$\frac{S^2}{\frac{d(N^2)}{dk}} = \frac{\pi}{2}\, n\, wt\, \frac{\left[(\bar{\beta}_x)^2 + (\bar{\beta}_z)^2\right]}{\overline{\beta^2}} \quad \frac{\tanh\frac{kt}{2}}{\frac{kt}{2}} \quad . \quad (54)$$

Thus, for large heads of large permeability, the narrow band S/N has been shown to be independent of the spacing, as was expected.

It is also to be noted that since

$$\int \overline{H_x}^2\, dx = \int \overline{C_x}^2\, dk$$

then

$$\int \overline{H_x}^2\, dx = \int \overline{H_z}^2\, dx \quad . \quad (55)$$

Of further interest is the signal-to-noise ratio in a given bandwidth. This is determined here for the zero-gap case. Since we have

$$d(N)^2 = n\, wt\, m^2\, \beta^2 \left[\frac{16\pi\, e^{-2kd}\left[1-e^{-2kt}\right]}{2kt}\right] dk \quad (56)$$

then

$$N^2 = n\, wt\, m^2\, \beta^2\, 16\pi\left[E_1\left(2k_1 d\right) - E_1\left(2k_1 (d+t)\right) - E_1\left(2k_2\, d\right) + E_1\left(2k_2 (d+t)\right)\right] \quad . \quad (57)$$

where $k_2 > k_1$.

For a bandwidth of $(k_2, k_1) = (0.2\pi, \frac{2\pi}{15})$ in mils^{-1} and a thickness t of 0.25 mil and 0.5 mil, we can evaluate S/N as a function of the spacing, d, where S here has most significance when taken either as d.c. or $\frac{k_2+k_1}{2}$. The noise power in the given bandwidth is shown in Figure 5 as a fraction of the noise power in infinite bandwidth.

6. Determination of the Means and Variances

It is now necessary to relate the variances, $\overline{\beta_x}^2$ and $\overline{\beta_z}^2$, and the means, $\bar{\beta}_x$ and $\bar{\beta}_z$, to the actual tape properties. This is merely a matter of calculation. By definition[3],

$$\overline{\beta_x}^2 = \int_{\Omega}\left[\cos(m,x) - \overline{\cos}(m,x)\right]^2 P_x(\theta,\phi)\frac{d\Omega}{4\pi}$$

$$\overline{\beta_z}^2 = \int_{\Omega}\left[\cos(m,z) - \overline{\cos}(m,z)\right]^2 P_z(\theta,\phi)\frac{d\Omega}{4\pi} \quad . \quad (58)$$

where Ω is a solid angle of integration. For no signal, $\overline{\cos}(m,x) = \overline{\cos}(m,z) = 0$. If the particle directions are randomly distributed, i.e., no particle orientation, then $P_x(\theta,\phi) = P_z(\theta,\phi) = 1$, and clearly,

$$\overline{\beta_x}^2 = \overline{\beta_z}^2 = \frac{1}{3}$$

For a saturated tape, we have:

$$\bar{\beta}_x = \int_{\Omega/z} \cos(m,x)\, P_x(\theta,\phi)\frac{d\Omega}{4\pi}$$

$$\bar{\beta}_z = \int_{\Omega/z} \cos(m,z)\, P_z(\theta,\phi)\frac{d\Omega}{4\pi}$$

Again, for no particle orientation, it is clear that

$$\bar{\beta}_x = \bar{\beta}_z = \frac{1}{2}$$

Thus, from Equations (27) and (26) we get:

$$S = 2\pi\,(n\, m\, wt) \quad . \quad (59a)$$

and

$$N^2 = 16\pi\left(\frac{2}{3}\right) n\, m^2\, wt\left[\frac{\lg(1+\frac{t}{d})}{t}\right] \quad . \quad (59b)$$

Therefore,

$$\frac{S^2}{N^2} = \frac{3\pi}{8}\, n\, wt\left[\frac{t}{\lg(1+\frac{t}{d})}\right] \quad . \quad (59c)$$

Thus, if $n = 10^{14}$ particles/cc, w = .25 inch, t = .5 mil, and $d = 50 \times 10^{-3}$ mils, we get:

$$S/N \cong 2.5 \times 10^7 = 80 \text{ db} \quad . \quad (60)$$

As stated previously, this result appears to be much too high if d.c. noise is to be explained. Thus, there are sources of noise in the tape not taken into account here. It is to be noted that if the head detected all components of m equally, then $\sigma^2(\beta) = 0$ and the noise would be zero.

7. Surface Noise

Implicit in our derivations so far has been the assumption that the magnetic surfaces of the head and tape

are perfectly smooth. This, of course, is never the situation. We now estimate the increase in noise due to a rough magnetic interface between tape and reproduce head.

The magnetic surface roughness is equivalent to a particle distribution function in the z direction, P(z), still assuming an otherwise homogeneous tape. The most that can be said about P(z) is that at z = d, P(z) has some value P_1; that at some slightly larger value, i.e., z = d + z_o, P(z) = 1; and that as z increases beyond z = d+z_o, P(z) = 1. In other words, the distribution of the magnetic sources is independent of z for z ≥ d + z_o. In order to get an estimate of the effect of magnetic surface roughness, P(z) is assumed to be a linear function of z. Therefore,

$$P(z) = z\left(\frac{1-P_1}{z_o}\right) + \frac{P_1\,(d+z_o) - d}{z_o} \qquad \ldots\ldots\ldots (61)$$

$$\int_d^{d+t} z^2\,P(z)\,dz \int_{-\infty}^{\infty} \frac{dx}{(x^2+z^2)^2} = \int_d^{d+t} \frac{P(z)}{z}\,dz$$

$$= \frac{\pi}{2}\left\{(1-P_1) + \left[\frac{P_1(d+z_o) - d}{z_o}\right] \lg\left(1+\frac{z_o}{d}\right) + \lg\left(\frac{1+\frac{t}{d}}{1+\frac{z_o}{d}}\right)\right\} \ldots\ldots\ldots (62a)$$

Thus, the ratio of the noise powers, with and without roughness, is:

$$\frac{N_1^2}{N^2} = 1 + \frac{\left(1-P_1\right)\left[1 - \frac{\lg\left(1+\frac{z_o}{d}\right)}{\frac{z_o}{d}}\right]}{\lg\left(1+\frac{t}{d}\right)} \qquad \ldots (63b)$$

The integral of the field is:

$$\int_d^{d+t} z\,P(z)\,dz \int_{-\infty}^{\infty} \frac{dx}{x^2+z^2}$$

$$= \frac{\pi}{2}\left[\int_d^{d+z_o} P(z)\,dz + \int_{d+z_o}^{d+t} dz\right]$$

$$= \frac{\pi}{2}\left[t - \left(\frac{1-P_1}{2}\right) z_o\right] \ldots\ldots\ldots (64a)$$

Then, the ratio of the saturated signals, with and without roughness, is:

$$\frac{S_1}{S_2} = 1 - \left(\frac{1-P_1}{2}\right)\frac{z_o}{t} \qquad \ldots\ldots\ldots (64b)$$

The relative S/N is:

$$\frac{\frac{S_1^2}{N_1^2}}{\frac{S^2}{N^2}} = \frac{\left[1 - \left(\frac{1-P_1}{2}\right)\frac{z_o}{t}\right]^2}{1 + \left(1-P_1\right)\left[\frac{\lg\left(1+\frac{z_o}{d}\right)}{\frac{z_o}{d}} - 1\right]} \Big/ \lg\left(1+\frac{t}{d}\right) \qquad \ldots (65a)$$

For the limiting case of extreme surface roughness, P_1 = 0. Then:

$$\frac{\frac{S_1^2}{N_1^2}}{\frac{S^2}{N^2}} = \frac{\left(1 - \frac{z_o}{2t}\right)^2}{1 + \left[\frac{1 - \frac{\lg\left(1+\frac{z_o}{d}\right)}{\frac{z_o}{d}}}{\lg\left(1+\frac{t}{d}\right)}\right]} \qquad \ldots . (65b)$$

We find only about 1 db increase for reasonable values of the parameters. For instance, we get a 1 db increase in S/N for the following case:

$$z_o = 50 \times 10^{-3} \text{ mils}$$

$$d = 5 \times 10^{-3}$$

$$t = .5$$

Thus, surface roughness contributes very little to the no-signal S/N. However, as will be shown in the section on a.c. noise, the amount of demodulated a.c. noise for recorded wavelengths of the order of magnitude of the surface roughness can be considerable.

8. D.C. Noise

Direct current noise is the transduced noise of a tape which is d.c.-magnetized to a given level. Experimentally, it is found that the tape noise rises sharply with the level of tape magnetization, reaching levels of 20 to 30 db (over the zero-signal noise) in the region of tape saturation. This increase in noise is not explainable in terms of the simple theory of non-interacting particles. Consideration of a d.c. magnetized collection of non-interacting particles will show that, on the contrary, the variance and thus the noise, should decrease. In fact for completely oriented particles, the noise should decrease to zero for complete magnetization. Thus, the effect of the interactions cannot be omitted in attempting to understand the magnitude of the increased noise accompanying magnetization in the tape.

That particle interactions cause the d.c. noise to be much larger than the zero-signal noise can be easily seen from the following argument. For zero magnetization, with or without particle interactions, the net magnetization is zero over any statistically large group of particles. In any such group there is a variance, which is the noise. If there is particle interaction, there is a correlation among the particles, especially among those closest to each other. The particles tend to orient such that their moments are in a direction opposite to the most strongly interacting particles, thus effectively decreasing the total number of magnetic sources or the average moment. For particles flipped by a d.c. field, the interaction field still tends to correlate them. However, instead of tending to line up at the axis around which they are disposed, any group of particles will tend to diverge from the axis, which will result in effectively less correlation among the particles and a considerable increase in the effective moments or number of magnetic sources. Therefore, clumping, which is a result of strong particle interaction (unless the clumps are larger than the statistical region) will cause the no-signal noise to decrease.

The exact analysis therefore involves particle interaction theory, a complicated subject that is not treated here. However, it is interesting (both in itself and as the basis for an analysis using particle interaction theory) to determine the d.c. S/N for the non-interaction theory. The prime concern here is to relate the magnetization to the number of flipped particles and thus to the d.c. noise. The total noise, which arises from both the unflipped particles (the source of the now decreasing zero-signal noise) and the flipped particles (the source of the d.c. noise) will increase. Another way of describing the same phenomena is to state that in the zero-signal case, the particles tend to line up in such a way as to limit the fluxpaths largely to themselves; whereas if there is any magnetization on the tape, the flux must, to some extent, leave the tape. It is seen, therefore, that d.c. noise is much larger than no-signal noise because the d.c. magnetization on the tape causes interacting particles, which form into effectively non-magnetic groups, to become effectively less interactive and thus to act as magnetic sources giving rise to both signal and noise. In no-signal noise, then, only a portion of the particles contribute to the noise. If the particles were sufficiently separated so that there would be little interaction, then the noise would increase to the maximum d.c. noise. Minimum no-signal noise is obtained, then, for maximum particle interaction, i.e. where the particles tend to form into rings. This, of course, would also tend to considerably increase the coercive force.

Thus:

$$N_T^2 = (1 - C)\, N^2 + C N_{d.c.}^2 \qquad \ldots\ldots (66)$$

where $2C$ is the fraction of all the particles that have flipped. For non-interacting particles, there is no distinction between the variances, and therefore no distinction between the noises of the flipped and unflipped particles. For interacting particles, of course, the noise is signal-dependent.

If a field is applied to the tape particles in a given direction, H_x, it will, upon reaching some given value, cause flipping of the particles in a cone $\frac{\pi}{4}$ radians from H_x[1]. As the field is increased, the width of the cone will increase symmetrically around the $\frac{\pi}{4}$ direction, approaching both the 0 and the π direction. If n is the particle density, then the density of randomly oriented particles having direction in a cone of angle $d\theta$ is:

$$n \sin\theta \, d\theta$$

as seen in Figure 7

The magnetization component, in the direction H_x, of particles in a cone of angular width $\pi - 2\theta_o$ is then:

$$M = n \int_{\theta_o}^{\pi/2 - \theta_o} m \cos\theta \sin\theta \, d\theta \qquad \ldots (67a)$$

and

$$M = \frac{m n}{2}\left[\cos^2\theta_o - \sin^2\theta_o\right] \quad \ldots\ldots\ldots (67b)$$

The density of the particles in the cone is

$$n_1 = n\int_{\theta_o}^{\pi/2-\theta_o} \sin\theta\, d\theta = n\left[\cos\theta_o - \sin\theta_o\right] \quad \ldots\ldots\ldots (68)$$

Therefore:

$$M = \frac{m}{\sqrt{2}}\, n_1\left[1 - \frac{1}{2}\left(\frac{n_1}{n}\right)^2\right]^{1/2} \quad \ldots (69a)$$

or

$$n_1 = n\left[1 - \left[1 - \left(\frac{2M}{nm}\right)^2\right]^{1/2}\right]^{1/2} \quad \ldots\ldots (69b)$$

For saturation, $n_1 = n$; then $M_o = \frac{nm}{2}$, as it should. Therefore, the ratio of d.c. noise to d.c. saturated noise is

$$\frac{N^2}{N_o^2} = \left[1 - \left[1 - \left(\frac{M}{M_o}\right)^2\right]^{1/2}\right]^{1/2} \quad \ldots\ldots (70)$$

Equation (70) is shown in Figure 6.

It is important to note that the magnetization is not proportional to the number of particles dc-magnetized, but that M increases with n_1 more slowly than linearly. Therefore, if the ratio of d.c. to noise level is taken, we have

$$\frac{S^2}{N^2} = n\,wt\,\frac{\left(\bar{\beta}_x \bar{G}_x\right)^2}{\overline{\beta^2}\;\overline{G_x^2}}\left[1 - \left[1 - \left(\frac{M}{M_o}\right)^2\right]^{1/2}\right]^{1/2} \quad \ldots\ldots (71a)$$

For $M << M_o$,

$$\frac{S^2}{N^2} = \frac{n\,wt}{\sqrt{2}}\left[\frac{\left(\bar{\beta}_x \bar{G}_x\right)^2}{\overline{\beta^2}\;\overline{G_x^2}}\right]\frac{S}{M_o} \quad \ldots\ldots (71b)$$

or

$$\frac{S}{N^2} = \sqrt{2}\; n\,wt\left[\frac{\left(\bar{\beta}_x \bar{G}_x\right)^2}{\overline{\beta^2}\;\overline{G_x^2}}\right] \quad \ldots\ldots\ldots (71c)$$

Or, from Equation (19a), if $\left(\frac{S^2}{N^2}\right)_o$ is the no-signal ratio, then

$$\frac{S}{N^2} = \left(\frac{S^2}{N^2}\right)_o \frac{1}{2\sqrt{2}\,M_o} = \left(\frac{S^2}{N^2}\right)_o \frac{1}{\sqrt{2}\,m\,n} \quad \ldots\ldots (71d)$$

For non-oriented, single-axis particle tapes, Equation (59c) states:

$$\left(\frac{S^2}{N^2}\right)_o = \frac{3\pi}{8}\left(\frac{n\,wt^2}{\lg\left(1 + \frac{t}{d}\right)}\right)$$

Then:

$$\frac{S}{N^2} = \frac{3\pi}{16\sqrt{2}\,M_o}\left(\frac{n\,wt^2}{\lg(1+\frac{t}{d})}\right) = \frac{3\pi}{8\sqrt{2}\,m}\left(\frac{wt^2}{\lg(1+\frac{t}{d})}\right) \quad \ldots\ldots (71e)$$

Thus, N is proportional to $\sqrt{S}$.

If the orientation is not random, then the density of the d.c. magnetized particles must be multiplied by the given density function, P (θ). Therefore, generally, for non-interacting acicular particles, we have:

$$S = n\int_{\theta_o}^{\pi/2-\theta_o} m_o \cos\theta \sin\theta\, P(\theta)\, d\theta \quad \ldots\ldots (72a)$$

$$n_1 = n\int_{\theta_o}^{\pi/2-\theta_o} \sin\theta\, P(\theta)\, d\theta \quad \ldots\ldots\ldots (72b)$$

The spectrum of d.c. noise for non-interacting particles, of course, will be the same as for the no-signal noise.

9. A.C. Noise

A.C. noise is simply a variable d.c. noise. Thus, for a perfectly homogeneous tape, its origin is the same as the origin of no-signal and d.c. noise. Here, however, as the signal level oscillates, the number of flipped particles, and thus the instantaneous noise level, also oscillates. The noise level, therefore, varies with the signal level. In order to determine the a.c. noise, it is necessary to go back to our original equations. Since n_1 now varies with x, it cannot be taken out of the integral defined by G. Therefore:

$$N^2 = wt\,m^2\left(\frac{1}{wt}\right)\int_v n_1(x)\left[\overline{\beta_x^2}\,\overline{H_x^2}(x,z) + \overline{\beta_z^2}\,\overline{H_z^2}(x,z)\right] dv \quad \ldots\ldots (73a)$$

and:

$$N^2 = n\,m^2\,wt\int_v \left\{1 - \left[1 - \left(\frac{M}{M_o}\cos k_o x\right)^2\right]^{\frac{1}{2}}\right\}^{\frac{1}{2}} \left[\overline{\beta_x^2}\,\overline{H_x^2}(x,z) + \overline{\beta_z^2}\,\overline{H_z^2}(x,z)\right] dv \quad (73b)$$

For a zero-gap head:

$$N_x^2 = n\,m^2\,wt\;\;\overline{\beta_x}^2\left(\frac{16}{t}\right)\int_d^{d+t} z^2\,dz \int_{-\infty}^{\infty} \frac{\left\{1-\left[1-\left(\frac{M}{M_o}\right)^2\cos^2 k_o x\right]^{1/2}\right\}^{1/2}}{(x^2+z^2)^2}\,dx \qquad \ldots\ldots (74a)$$

$$N_z^2 = n\,m^2\,wt\;\;\overline{\beta_z}^2\left(\frac{16}{t}\right)\int_d^{d+t} dz \int_{-\infty}^{\infty} \frac{x^2\left\{1-\left[1-\left(\frac{M}{M_o}\right)^2\cos^2 k_o x\right]^{1/2}\right\}^{1/2}}{(x^2+z^2)^2}\,dx \qquad \ldots\ldots (74b)$$

For small magnetization ($\frac{M}{M_o} << 1$) we have:

$$N_x^2 \cong n\,m^2\,wt\;\;\overline{\beta_x}^2\sqrt{2}\left(\frac{16}{t}\right)\frac{M}{M_o}\int_d^{d+t} z^2\,dz \int_0^{\infty} \frac{|\cos k_o x|}{(x^2+z^2)^2}\,dx \qquad \ldots\ldots (75a)$$

$$N_z^2 \cong n\,m^2\,wt\;\;\overline{\beta_z}^2\sqrt{2}\left(\frac{16}{t}\right)\frac{M}{M_o}\int_d^{d+t} dz \int_0^{\infty} \frac{x^2|\cos k_o x|}{(x^2+z^2)^2}\,dx \qquad \ldots\ldots (75b)$$

The integrals and their Fourier transforms in these expressions for noise probably cannot be evaluated in a closed form. Therefore, we consider a.c. bias produced not by a magnetization M cos $k_o x$, but rather by M cos^2 $k_o x$, giving us the easily integrable noise components:

$$N_x^2 = K_x\int_d^{d+t} z^2\,dz \int_0^{\infty} \frac{\cos^2 k_o x}{(x^2+z^2)^2}\,dx \qquad \ldots\ldots (76a)$$

$$N_z^2 = K_z\int_d^{d+t} dz \int_0^{\infty} \frac{x^2\cos^2 k_o x}{(x^2+z^2)^2}\,dx \qquad \ldots\ldots (76b)$$

where

$$K_x \equiv n\,m^2\,wt\;\;\overline{\beta_x}^2\sqrt{2}\left(\frac{16}{t}\right)\frac{M}{M_o}$$

There will be two differences in the noise spectra for the two cases: (1) For the signal M cos $k_o x$, the noise will be modulated largely by a wavenumber $\frac{k_o}{2}$; while for the signal M cos^2 $k_o x$, the modulation will be by the wavenumber k_o. (2) For the signal M cos $k_o x$, there will be modulation by larger wavenumbers; while for the signal M cos^2 $k_o x$, the modulating wavenumber will be only k_o. These differences can be seen clearly by drawing the curves for $|\cos k_o x|$ and cos^2 $k_o x$. Thus, it is possible to approximate the a.c. noise for a signal M_o cos $k_o x$ by evaluating the noise for a signal M cos^2 $k_o x$. Then:

$$\int_0^{\infty} \frac{\cos^2(k_o x)}{(x^2+z^2)^2}\,dx = \frac{1}{2}\left\{\int_0^{\infty} \frac{\cos(2k_o x)}{(x^2+z^2)^2}\,dx + \int_0^{\infty} \frac{dx}{(x^2+z^2)^2}\right\}$$

$$= \frac{\pi}{8z^3}\left[+e^{-2k_o z}(1+2k_o z) - 1\right] \qquad \ldots\ldots (77a)$$

Also:

$$\int_0^{\infty} \frac{x^2\cos^2(k_o x)}{(x^2+z^2)^2}\,dx = \frac{1}{2}\left\{\int_0^{\infty} \frac{x^2\cos(2k_o x)}{(x^2+z^2)^2}\,dx + \int_0^{\infty} \frac{x^2\,dx}{(x^2+z^2)^2}\right\} = \frac{\pi}{8z}\left[e^{-2k_o z}(1-2k_o z) + 1\right] \qquad \ldots\ldots (77b)$$

Thus,

$$N^2 = N_x^2 + N_z^2 = \frac{\pi K}{4t}\int_d^{d+t} \frac{\left|1+e^{-2k_o z}\right|}{z}\,dz$$

and

$$N^2 = \frac{\pi K}{4t}\left[\lg\left(1+\frac{t}{d}\right) + E_1(2k_o d) - E_1\left[2k_o(d+t)\right]\right] \qquad \ldots\ldots (78)$$

As expected, it is seen from Equation (78) that the noise consists of two parts: (1) d.c. noise, one-half the magnitude of which arises from a d.c. tape magnetization; (2) a.c. noise, whose magnitude varies with the impressed frequency. As $k_o \longrightarrow \infty$, the a.c. component decreases to zero. As $k_o \longrightarrow o$, the a.c. component approaches the d.c. component. Thus, as $k_o \longrightarrow \infty$, the total noise drops to one-half the noise produced by a d.c. signal of the same amplitude.

It is now of particular interest to determine the spectrum of this a.c. noise. Since

$$\int_o^\infty g^2(k,z)\,dk = \int_o^\infty f^2(x,z)\,dx$$

where

$$g(k,z) = \sqrt{\frac{2}{\pi}} \int_o^\infty f(x,z)\cos kx\,dx$$

we have:

$$g_1(k,z) = \sqrt{\frac{2}{\pi}} \int_o^\infty \frac{\cos k_o x \cos kx\,dx}{(x^2+z^2)}$$

$$= \frac{1}{2}\sqrt{\frac{2}{\pi}} \int_o^\infty \left[\frac{\cos(k_o+k) + \cos(k_o-k)}{(x^2+z^2)}\right] dk$$

$$= \frac{1}{2}\sqrt{\frac{\pi}{2}} \left[\frac{e^{-(k_o+k)z} + e^{-|k_o-k|z}}{z}\right] \quad \ldots\ldots (79a)$$

Then:

$$g_1^2(k,z) = \frac{\pi}{8z^2} \begin{cases} e^{-2k_o z}\left[e^{-2kz} + e^{2kz} + 2\right] & k \le k_o \\ e^{-2kz}\left[e^{-2k_o z} + e^{2k_o z} + 2\right] & k \ge k_o \end{cases}$$

$\ldots\ldots\ldots$ (79b)

Then:

$$\int_d^{d+t} z^2\,dz \int_o^\infty \frac{\cos k_o x}{(x^2+z^2)^2}\,dx$$

$$= \frac{\pi}{8}\int_d^{d+t} \begin{cases} e^{-2k_o z}\left[e^{-2kz} + e^{2kz} + 2\right] & k \le k_o \\ e^{-2kz}\left[e^{-2k_o z} + e^{2k_o z} + 2\right] & k \ge k_o \end{cases} dz$$

$$= \frac{\pi}{8} \left\{ \frac{e^{-2(k_o+k)d}\left[1 - e^{-2(k_o+k)t}\right]}{2(k_o+k)} + \frac{e^{-2|k_o-k|d}\left[1 - e^{-2|k_o-k|t}\right]}{2|k_o-k|} + \begin{cases} \dfrac{e^{-2k_o d}\left[1 - e^{-k_o t}\right]}{2k_o} & k \le k_o \\ \dfrac{e^{-2kd}\left[1 - e^{-k_o t}\right]}{2k} & k \ge k_o \end{cases} \right\} \quad \ldots\ldots (80)$$

The evaluation of the integral for N_z^2 results in an extremely complicated expression. Since its spectrum is about the same as that for N_x^2, it will not be calculated.

We see that the narrow bandwidth noise is

$$d(N_x^2) = \frac{K\pi}{8} \left\{ \frac{e^{-2(k_o+k)d}\left[1 - e^{-2(k_o+k)t}\right]}{2(k_o+k)} + \frac{e^{-2|k_o-k|d}\left[1 - e^{-2|k_o-k|t}\right]}{2|k_o-k|} + \begin{cases} \dfrac{e^{-2k_o d}\left[1 - e^{-2k_o t}\right]}{2k_o} & k \le k_o \\ \dfrac{e^{-2kd}\left[1 - e^{-2kt}\right]}{2k} & k \ge k_o \end{cases} \right\} dk$$

$\ldots\ldots$ (81a)

The noise resulting from the a.c. magnetization of the tape can be viewed as a modulation of the carrier by the Fourier components of the d.c. noise. Therefore, we can write the noise in terms of the Fourier components of the d.c. noise $k_1 \equiv |k_o - k|$:

$$d(N_x^2) = \mp \frac{K\pi}{8} \left\{ \frac{e^{-2k_1 d}\left[1 - e^{-2k_1 t}\right]}{2k_1} \right.$$

$$+ \left\{ \frac{e^{-2(2k_o - k_1)d}\left[1 - e^{-2(2k_o - k_1)t}\right]}{2(2k_o - k_1)} \right.$$

$$\left. \left. + \frac{e^{-2(2k_o + k_1)d}\left[1 - e^{-2(2k_o + k_1)t}\right]}{2(2k_o + k_1)} \right\} \right.$$

$$\left. + \frac{e^{-2k_o d}\left[1 - e^{-2k_o t}\right]}{2k_o} \right.$$

$$\left. + \frac{e^{-2(k_1 + k_o)d}\left[1 - e^{-2(k_1 + k_o)t}\right]}{2(k_1 + k_o)} \right\} dk_1 \qquad \begin{matrix} k_1 \leq k_o \\ k_1 \geq k_o \end{matrix}$$

. . . . . (81b)

It is seen that if $k_o \longrightarrow \infty$, then the a.c. noise spectrum approaches the d.c. noise spectrum. This agrees with the broadband calculation for a.c. noise. Thus, the a.c. noise arising from the tape is truely a modulation noise, even with the assumption that the tape is perfectly homogeneous.

If the irregularities in the tape, such as surface roughness, are taken into account, then as $k_o \longrightarrow \infty$, we do not expect the a.c. noise to approach the d.c. noise. Thus, the percentage modulation of the transduced a.c. recorded signal, due to the roughness of the tape surface acting as a variable head-tape spacing, increases as $k_o \longrightarrow \infty$. Therefore, the demodulated noise at a given k_1 will increase as $k_o \longrightarrow \infty$[9]. Any irregularities in the tape will add significantly to the demodulated noise as $k_o \longrightarrow \infty$.

Notation

d	= head-tape spacing
E (b)	= expectation of b
g	= Green's function for the flux
H_x, H_z	= magnetic field components
$H_x(x_1)$, $H_z(x_2)$	= maximum H_x and H_z, respectively
k	= $\frac{2\pi}{\lambda}$, the wave number
k_o	= the recorded wave number
M	= tape magnetization
m	= magnetic moment of the tape particle
m_x, m_z	= x and z components, respectively, of the tape particle magnetic moment
m_o	= component of tape particle magnetic moment producing head-detected flux
N	= noise from the tape transduced by the head
n	= tape particle density
n_1	= effective tape particle noise source density
r_x, r_z	= effective lengths of tape detected by the head in a noise determination
$\frac{dr_x}{dk}$, $\frac{dr_z}{dk}$	= noise spectrum resolution factors
R_x, R_z	= effective lengths of tape detected by the head in a signal determination
S	= signal from reproduce head, flux or output voltage
t	= thickness of tape
w	= width of tape
β_x, cos (m,x) β_z, cos (m,z)	= the components of the particle moments in the x and z directions, respectively
λ	= wavelength
$\sigma^2(b)$, $\overline{\beta^2}(b)$	= standard deviation of b
ϕ	= head-detected magnetic flux from the tape

References

1. Stoner, E.C., and E.P. Wohlfarth, "Mechanism of Magnetic Hysteresis and Heterogeneous Alloys", Phil. Trans. Royal Soc., A 240, 1948, p 599.

2. Westmijze, W.K., Philips Research Reports, vol 8, 1953, Research Laboratory of N.V. Philips, Eindhoven, Netherlands.

3. Feller, W., An Introduction to Probability Theory and Its Applications, vol 1, 2nd ed., John Wiley & Sons, Inc., 1957.

Note: or any other book on statistics or probability.

4. Mann, P.A., Archiv der Elektrischen Ubertragung, vol 11, 1957, p 97.

5. Daniel, E.D., Ampex Research Report AEL 1, Ampex Corporation, Dec 1960.

6. Morse, P.M., and H. Feshbach, Methods of Theoretical Physics, vol 1, McGraw-Hill Book Company, Inc., 1953.

7. Smaller, P., Proceedings of Magnetic Recording Technical Meeting of Oct 4,5, 1956, Bulletin No. 94, Armour Research Foundation, 1956.

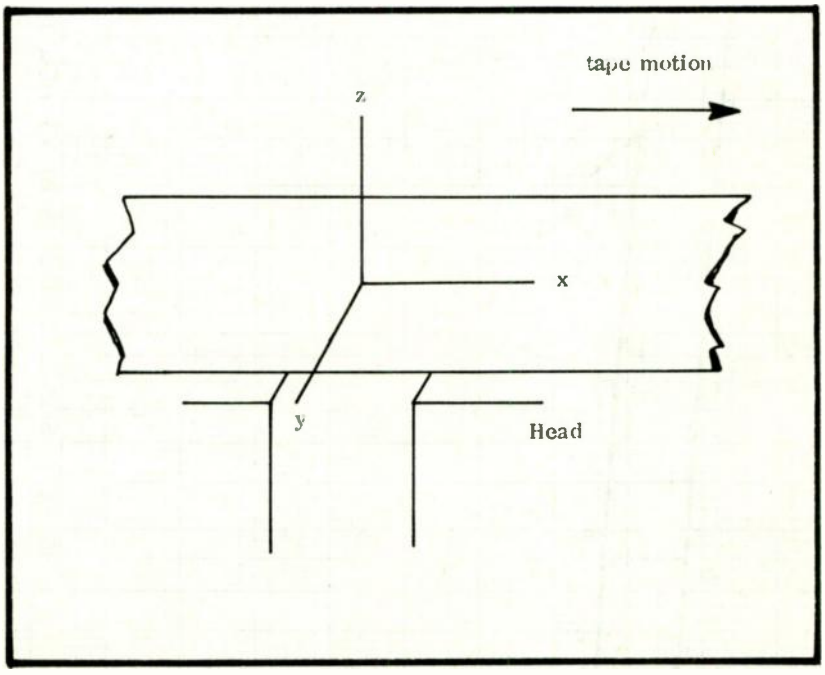


Fig. 1. Tape-head coordinate system.

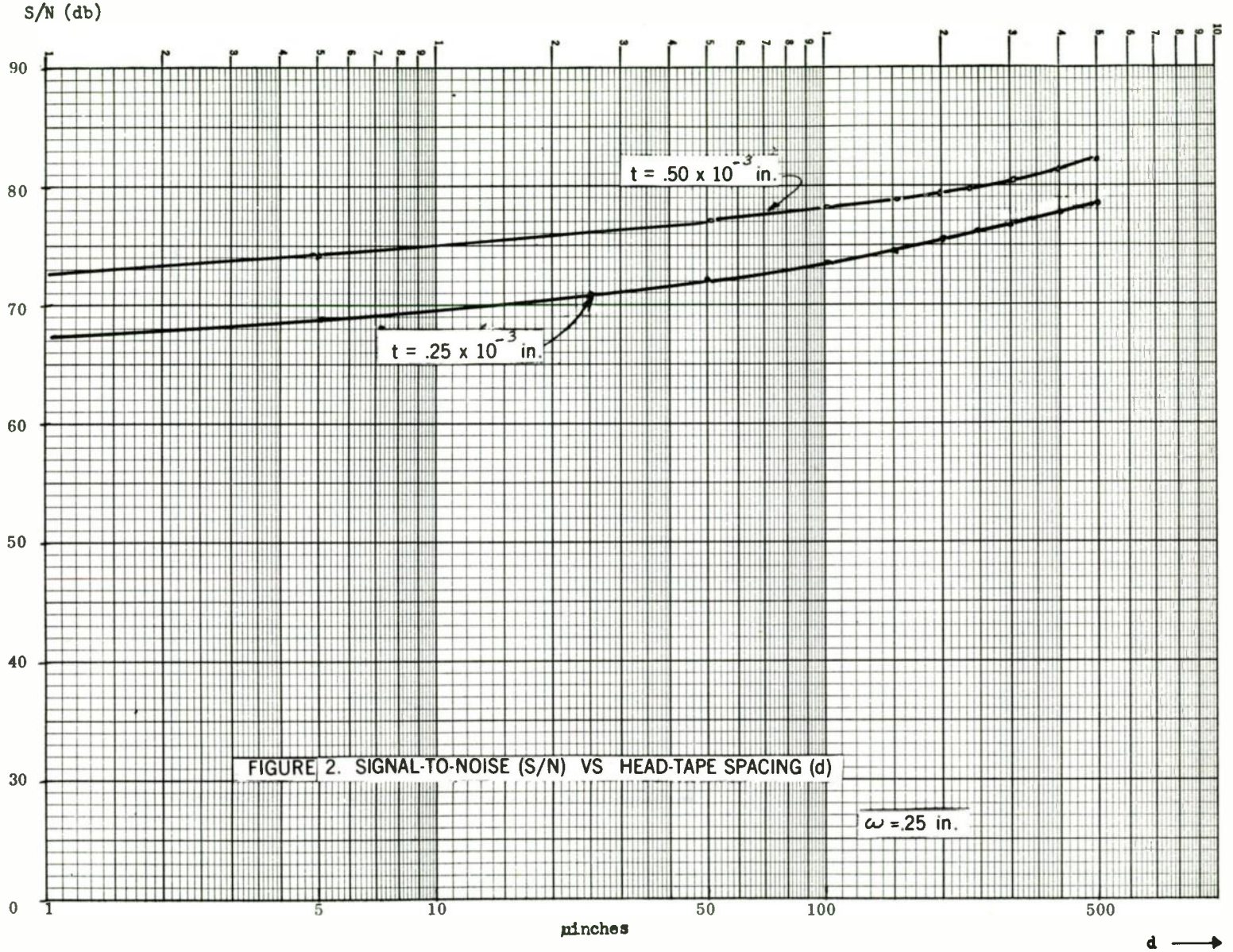


FIGURE 2. SIGNAL-TO-NOISE (S/N) VS HEAD-TAPE SPACING (d)

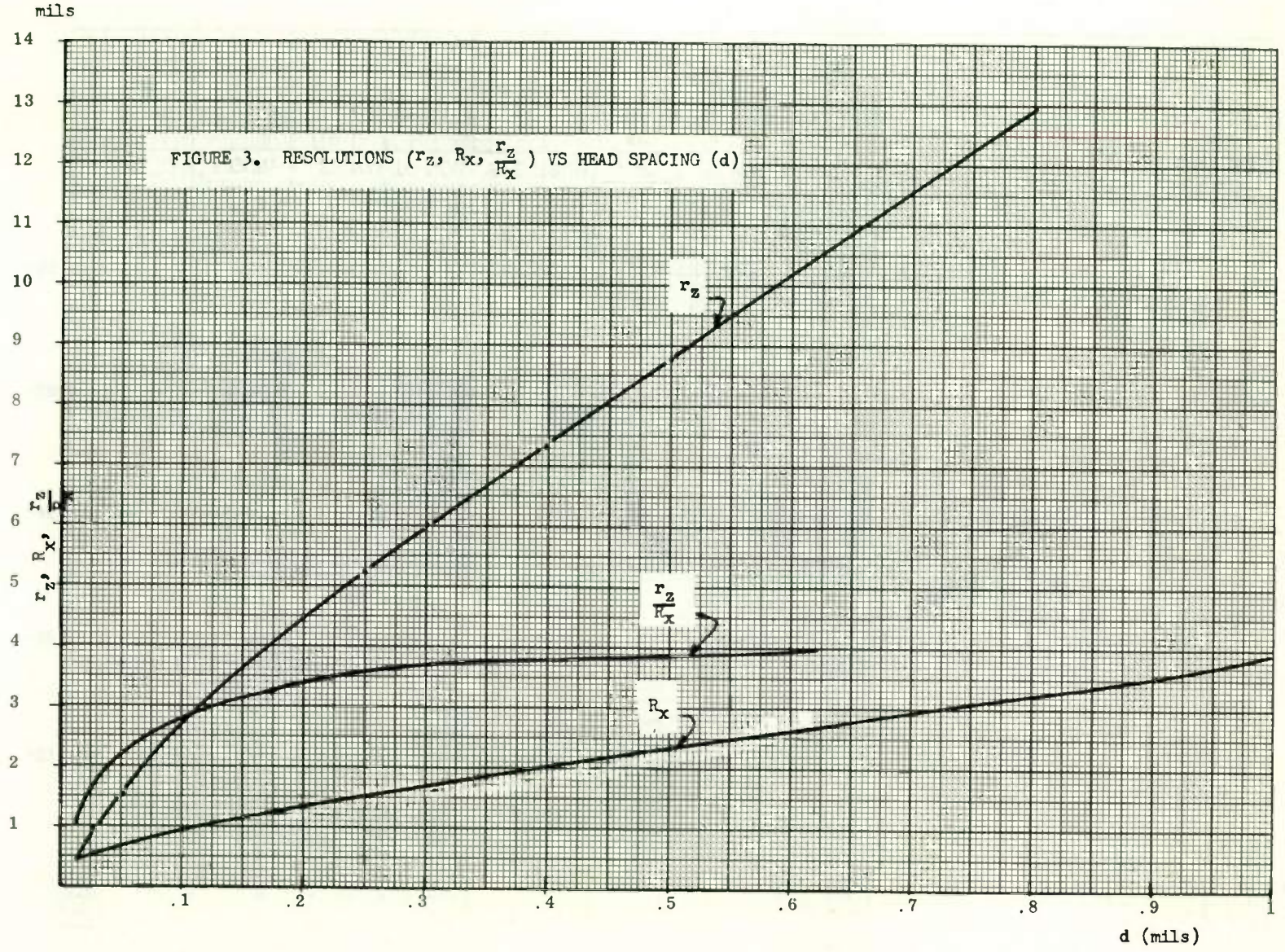


FIGURE 3. RESOLUTIONS (r_z, R_x, $\frac{r_z}{R_x}$) VS HEAD SPACING (d)

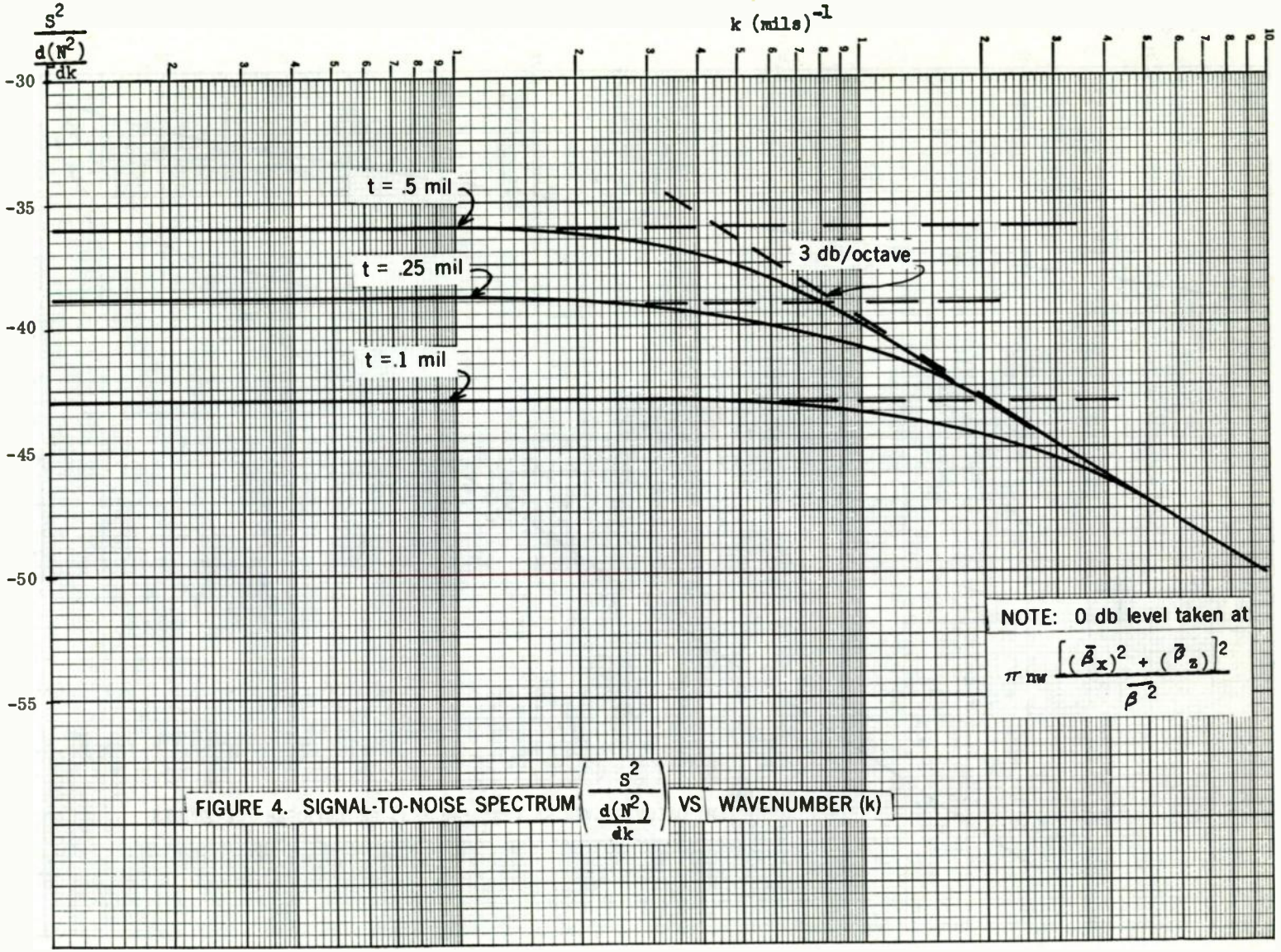


FIGURE 4. SIGNAL-TO-NOISE SPECTRUM $\left(\frac{S^2}{\frac{d(N^2)}{dk}}\right)$ VS WAVENUMBER (k)

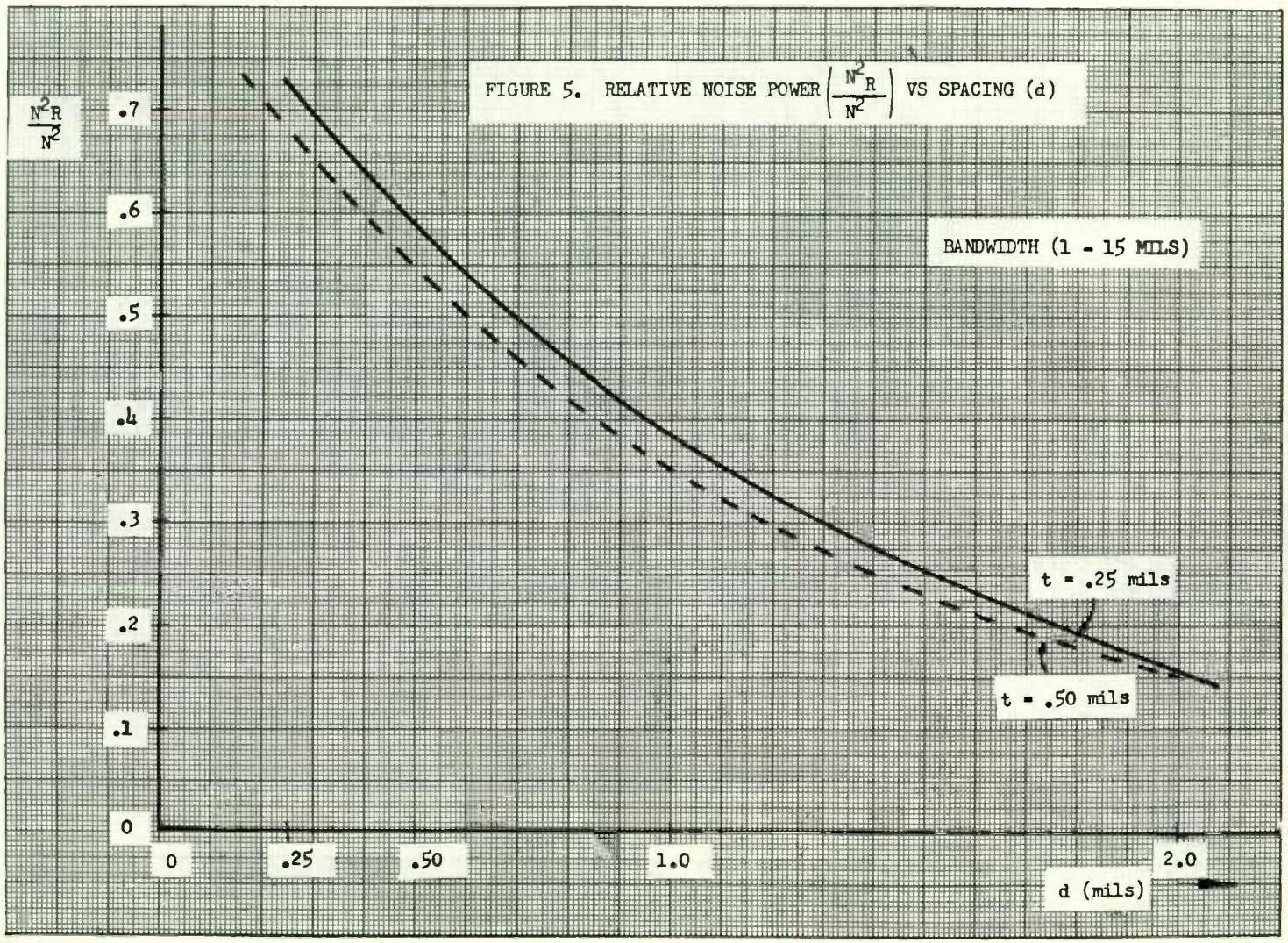


FIGURE 5. RELATIVE NOISE POWER $\left(\frac{N^2_R}{N^2}\right)$ VS SPACING (d)
$\frac{N^2_R}{N^2}$
.7
.6
.5
.4
.3
.2
.1
0
BANDWIDTH (1 - 15 MILS)
t = .25 mils
t = .50 mils
0
.25
.50
1.0
2.0
d (mils)

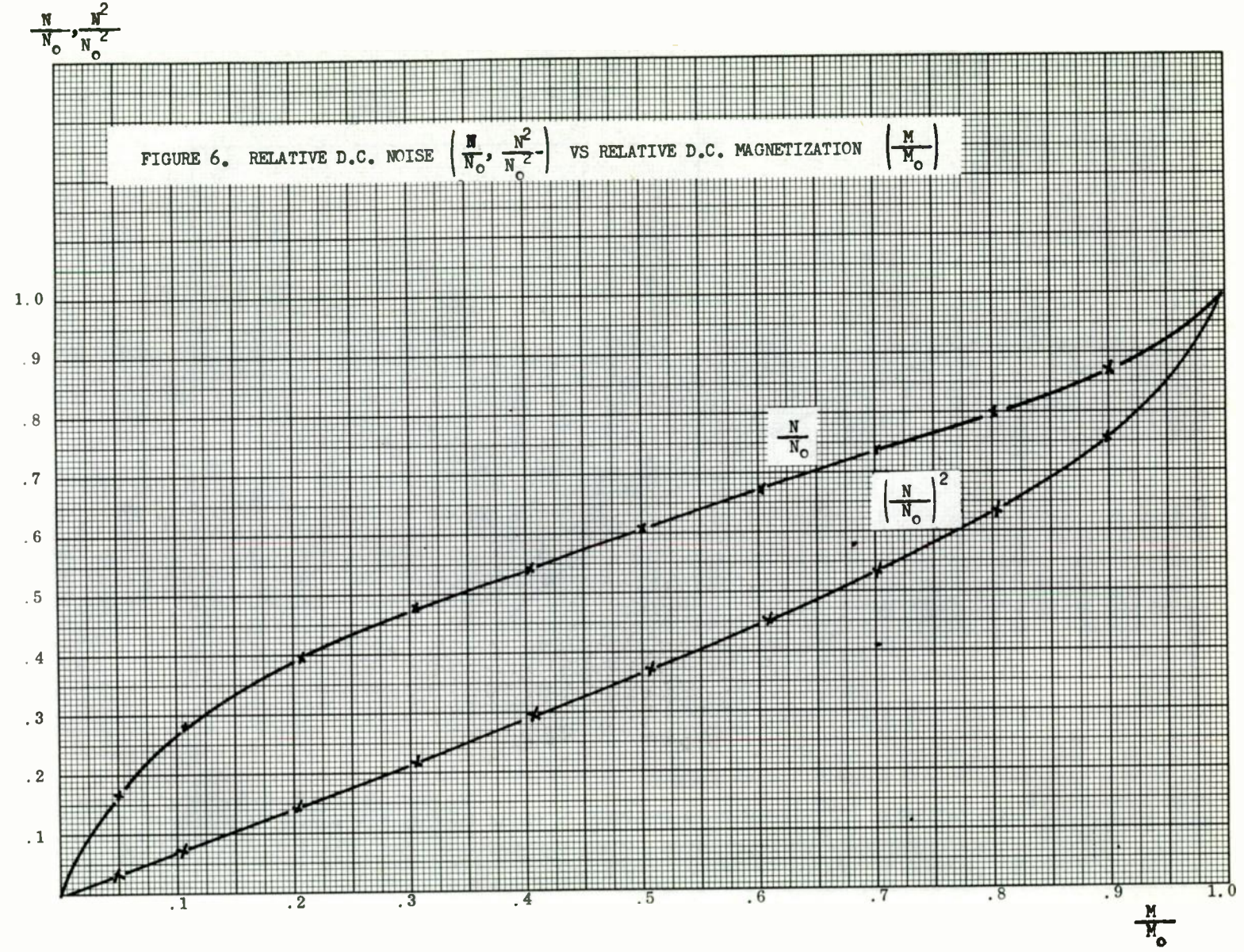


FIGURE 6. RELATIVE D.C. NOISE $\left(\frac{N}{N_o}, \frac{N^2}{N_o^2}\right)$ VS RELATIVE D.C. MAGNETIZATION $\left(\frac{M}{M_o}\right)$

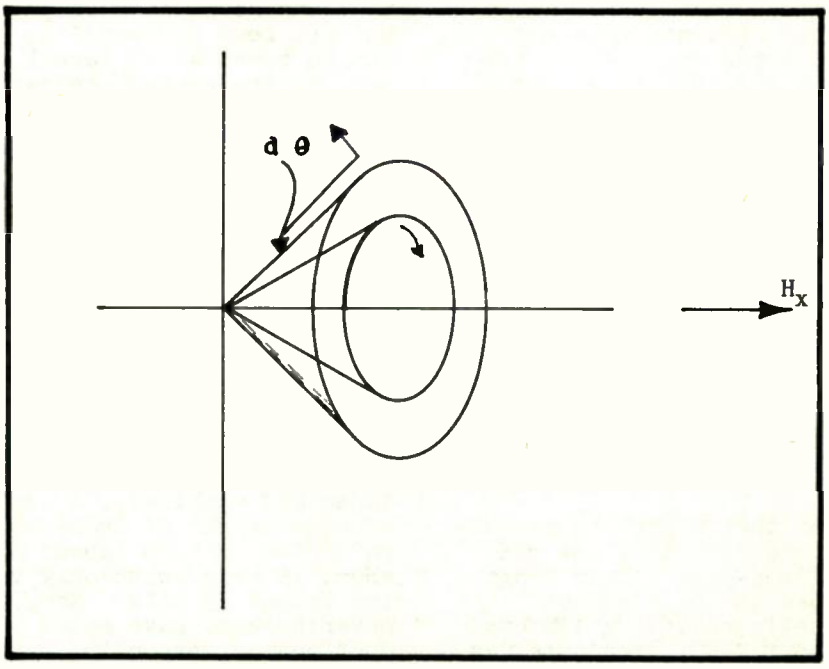


Fig. 7. Region of particle flipping.

DROP-OUTS IN INSTRUMENTATION MAGNETIC TAPE RECORDING SYSTEMS

Robert H. Carson
U. S. Naval Research Laboratory
Washington, D. C.

A method of measuring drop-outs from magnetic tape recording systems has been designed and built so that quantitative data can be quickly obtained from any tape on any machine and under any condition. Conclusions are made as to the relative effects of tape, machines, reels, recording processes, reproducing processes, environmental dirt, tensions, etc on the drop-out count. It has been found that drop-out measurements can be used to evaluate tape and machines in regards to wear characteristics, to eyaluate tape cleaning methods, and to determine techniques in handling tape and operating tape machines to minimize drop-out effects both for analog and digital applications.

Introduction

Background

The phenomena that occurs in an output signal of a tape recorder, now generally known as "drop-outs", have long been known by those who have worked closely with magnetic recording systems. However, up to about 1950, drop-outs, as known today, were very numerous and were usually masked by an overall variation in the output level that was continuously changing. The envelope from such a tape is shown in Fig. 1A. A constant level sine wave signal was recorded with a recorded wave length of 1 mil.

Due to steady improvement in both tapes and machines, the output of present-day systems may be as shown in Fig. 1B. Unfortunately, however, this is an exception rather than the rule. The usual envelope is the one shown in Fig. 1C, though this represents the output of a system somewhat worse than the average.

These discrete drops in signal are known as drop-outs. With an envelope as in 1A the modulation noise figure was a good indication of the envelope character, but with discrete drops in signal, this figure has little significant meaning. Various shapes of drop-outs are shown in Fig. 2. These were selected from a large number of tapes.

Definition

The definition of the term drop-out has not been standardized so its meaning can be somewhat different as used by various groups who use magnetic recording systems.

In this study, a drop-out has been considered any reduction of signal level, as reproduced from a magnetic recording system, of a specified amount (either in db or per cent) and lasting no longer than 100 milliseconds at a tape speed of 7-1/2 ips (12-1/2 ms at 60 ips) provided a constant level signal was recorded. The time indicated represents 75 mils in actual length of tape.

This definition resulted from study and experience with many magnetic tapes. In making a thorough search with many tapes and machines, it was found that the average length of drop-out with a signal reduction of 3 db (about 30 per cent) was about 15 mils, with only very few extending beyond 75 mils. Many observers, nevertheless, have noted that signal reductions corresponding to much longer lengths of tape - even up to several inches - may be found. It is true that such reductions can be found under certain special conditions, but they do not result from the same causes as the shorter drop-outs.

Drop-outs that fit the definition result almost entirely from loss of contact due to nodules or holes in the tape surface, or from foreign particles either embedded in the tape or on the surface or a combination of both. Longer drops in signal level are caused by skew (azimuth misalignment) or loss of contact over longer lengths of tape due to physical deformities of the tape that may result from improper tensions applied by machine and reels and from severe temperature and humidity changes.

Though these longer drops in output can be quite important, they were not included in this study since it was considered necessary to first determine the basic non-uniformity characteristics of

tapes and machines under normal laboratory conditions. This study was initiated because of the many contradictory opinions generally prevalent concerning which tapes have the most or least drop-out effect, which machines causes more or less drop-outs, and to what extent various environmental conditions affects the signal loss from drop-outs.

Measuring Technique

The design, construction, and calibration of the measuring system was a major problem and the final design resulted from considerable trial and error experimental experience. These design problems will not be detailed, but to aid in the interpretation of the data, some understanding of the basic measuring technique is necessary.

Many systems using demodulation techniques have been used to show the non-uniformity of output, and for modulation envelopes that were continuously changing, such systems can give a general figure of merit. With only discrete drops in output, a more precise system must be utilized. Since the objective was to make a basic study of factors affecting drop-out formation and not just total loss of information, a digital system was constructed that would count the number of times that the signal drops below a selected threshold level, and in addition, would also measure the length of time that the signal stays below the threshold.

Fig. 3 is a block diagram of the basic system shown for one threshold level. The system shown is the latest revision of the measuring system. It is completely solid state, completely digital, and will operate at frequencies up to 100 kc. The system used previously was all vacuum tube plug-in units, was limited to frequencies up to 15 kc, and used one shot controlled gates that measured the time that the signal stayed below the threshold level. In the present system, the timing system is based on a free running multivibrator, synchronized to the signal frequency, whose output is fed into a binary counter which can be tapped off at any count to control a gate through which the counting pulse must pass. This system inspects every cycle or pulse of the reproduced signal and if the signal is above the threshold, the one shot will fire. The leading edge of the one shot pulse is the counting pulse and will go through any gates that are open. The trailing edge of the one shot pulse is used to reset all flip-flops so that the gates are closed, if not already closed.

All pulses going through the gates are counted in decimal counters and after the run is completed, the various numbers are either recorded on a data sheet or punched out on tape for entry into a computer.

In addition, a 20 channel event recorder is available so that the counting pulses can also be made to operate the marking pen. This gives a graphic presentation of the location of drop-outs along the tape.

A timing system is used that allows a tape to be rerun and started at exactly the same place. Therefore, a complete history of any particular drop-out can be followed through any number of runs.

Results

General Survey

In starting to investigate any complex phenomena, the experimenter is confronted with decisions as to what factors to investigate first. In this study, it was thought advisable to first make a general survey of the types of tapes available to determine the general distribution of drop-outs, in amplitude and in length.

Amplitude Distribution. Fig. 4 is a plot of the number of drop-outs per 1000 ft of tape versus per cent signal reduction, regardless of length. While these curves are from 1959 tape, data from recent tapes show the same general curve shape, but may vary in actual numbers. The curve D-4 shows why it is often difficult to obtain consistent measurements from a tape, even when run through the same machine on successive passes. As shown, if the output level changes from one run to the next by as little as 5% in the 3 db threshold region, the actual count of drop-outs could vary about 250 out of a total of 800. A 5% change in output level can easily occur due to normal tension variations in the best instrumentation recorders.

In order to obtain more consistent

data, it was decided that on most tests, an AVC circuit would be used to drive the drop-out measuring equipment. A time constant was selected that would allow all drop-outs to pass but would eliminate changes of the average output level due to machine tension, tape skew and other long term effects.

Length Distribution. This general study also showed that the distribution of drop-outs by length was very similar for all standard tapes. This is shown in Fig. 5. The tape that is somewhat different is one that has a thin plastic coating over the oxide.

Data Requirements

Using this general survey as a basis, a system was selected that would give the most usuable information with the least possible data numbers, and also considering the practical limitation of the number of thresholds and gate lengths that could be used simultaneously. During a period of two years, over 185 reels of instrumentation tape representing 38 types from four manufacturers have been tested using 3 amplitude thresholds, 3, 6, and 12 db, and 4 time gates for each threshold, 3-8, 9-15, 16-35, and over 36 cycles. As a result of considerable effort spent in determining the most efficient data processing technique and data presentation, it now seems possible to obtain all usuable information with only 9 data numbers per run. These include the total count for 5 amplitude thresholds, 2, 3, 6, 12, and 18 db, with the 4 time gates for the 3 db threshold only. It is believed that this amount of data from each run will make possible a thorough evaluation of tapes and machines in respect to any specified application. There is a strong possibility that with added experience and knowledge, even less data may be required for such evaluations. Undoubtedly the ultimate objective would be to arrive at an index figure of merit for certain major types of application such as an analog index, digital index, etc.

Data Processing

Even after very few tapes had been run and drop-outs counted, it was apparent that the processing of the data into a useful form was going to be a difficult problem. It seemed that the only solution would be the use of a computer. Considerable time and effort was spent in determining the way in which the data was to be processed and the format in which to present it. It was realized that a true statistical analysis could not be made, but it did seem that statistical methods of treatment of the data could be utilized.

Data Presentation Format. Several formats of data tables were programmed through a computer and the final results submitted to various laboratory personnel and visitors for comment. As a result a compact format was selected that is believed to contain all the information required to evaluate a tape and/or machine. Table 1 represents the data from testing of any number of reels of tape of the same type, run on the same machine, and represents the initial condition of the tape as received from the manufacturer. A brief explanation of the table will be given.

1. The left column "Amp. Thres." designates the amplitude threshold used.

2. The column "Total Number" first gives the average number of all reels and runs, and below this, the standard deviation of the average from reel to reel. This shows the consistency from reel to reel.

3. The next column "Reel Std. Dev." shows first the average of the standard deviation of the means for each reel, and the lower figure shows the standard deviation of this mean of standard deviations. The top figure indicates how much variation to expect within any one reel of this type of tape while the lower figure indicates the consistency of the standard deviation from one reel to another.

4. The figures in the "Slope Column" indicate the average per cent increase (positive number) or decrease (neg. number) in drop-outs which can be expected on each successive pass of the tape relative to the first pass.

5. The column "Minimum Maximum" shows the minimum and maximum drop-out counts that were encountered in the complete test.

6. The next column shows the percentage of total drop-outs represented by the drop-outs for the particular

amplitude threshold shown.

7. The rest of the data is concerned with the breakdown of the drop-outs into lengths, with a percentage of the total number and its standard deviation shown for each length. This data is useful in determining, for any given application, how much information would be lost. It has been found that only the 3 db figures are sufficient. Therefore all future data tables will have only the 3 db drop-outs broken down into lengths.

This data can be arranged in graphs and tables of various types to bring out certain relationships, such as to determine if there is any significant difference in backing material among all manufacturers, or whether all tape of the same manufacturer exhibit similar characteristics.

Data Graphs. One type of plot was made from all this data and is shown in Figs. 6 and 7. A standard oxide coated on 1-1/2 mil acetate, 1-1/2 mil Mylar and 1 mil Mylar was selected from each manufacturer, where possible, and the average number of 3 db drop-outs per 1000 ft, the standard deviation reel to reel, and the average standard deviation of all reels are shown for tapes of 1959 and of 1960 production. This same type of information will be available this summer for 1962 production tape.

Wear Tests. After running very few tapes, it seemed that the measurement of drop-outs could be a very sensitive method of determining the wear characteristics of a tape with only a few passes. Therefore tests were set up to run every tape several times with a new recording made on each run since evidence clearly shows that the recording process determines the drop-out condition of a tape to a greater extent than the reproducing process. This had been assumed for years as a result of obtaining considerable qualitative data, but the use of the chart recorder has given quantitative data to prove it.

Drop-out Repeatability. Table 2 shows the per cent repeatability of drop-outs on reproducing runs only, and then on recording-reproducing runs. Chart recordings were used and drop-outs were compared each to each.

As can be seen from the table, for reproducing runs only 70 per cent of the drop-outs that occurred once, reappeared on all runs. Only 10% appeared only on one run. However, with new recordings made for each run, only 45% of the drop-outs appeared on each run, but 20% appeared only once. It should be pointed out that these figures show the most favorable relation since all runs were made under fairly high bias conditions. If the bias were to be adjusted for maximum output, as is usual, the repeatability of drop-outs from one recording to another would drop considerably. And in non-biased saturated systems, the repeatability from one run to another can be so low that it is difficult to obtain any kind of statistical correlation.

Figs. 8 and 9 are short samples of the chart recording that was used to obtain the repeatability figures of Table 2. Fig. 8 shows a tape that was reproduced 4 times. There is a slight timing error between runs, but it is always easy to recognize and compensate for such error. The chart paper is generally run at 3 inches per minute and only on very bad tape do drop-outs occur so frequently that one mark on the chart may represent more than one drop-out. The tape speed was 7-1/2" ips. Since one minute of time represents 37.5 ft of tape, the chart is quite compressed. Higher chart speed (up to 12 ipm) have been used to study a particular phenomenon.

A study of the 6 db and 12 db drop-outs shows that generally they are quite repeatable, and even when not, a drop-out of lesser magnitude can usually be found to correspond to the non-repeating higher magnitude drop-out. Two of these cases are noted in the 6 db plot. The 6 db drop-out on the first reproduction at the far left (A) did not repeat. It can be seen, however, that although a 3 db drop-out occurred on both the first and second runs, it did not appear on the third and fourth runs.

There is a different situation with the two close 6 db drop-outs in the fourth run shown at C. These did not show up until the fourth run, but an inspection of the 3 db charts show that these drop-outs were there all the time. A final example is that of the 12 db

drop-out at E. It appeared only on the third run, while a 3 db and 6 db drop-out occurred at this point for each run. This is an indication of how machine variation in tension can cause such different drop-out counts on successive runs.

Fig. 9 shows another chart sample, one made where each new run was of a new recording. While there still are a large number of drop-outs that repeat exactly, the general pattern does appear more irregular than Fig. 8. Some specific examples are noted. The 6 db drop-out at A, and the corresponding 3 db drop-out at B, appeared only on the fourth recording. There were 6 db drop-outs, C and D, on recordings 1 and 4, with a 12 db drop-out, E, appearing only on the 4th recording. There were 3 db drop-outs at this location on all recordings.

Another general effect can be seen from these charts, although on such a short sample, it is not so striking. Drop-outs often occur in groups. A section of tape that contains large drop-outs often is preceded and followed by smaller drop-outs. This can be seen in Fig. 9 at the left of the chart.

A brief explanation of why the recording process is more responsible for the formation of drop-outs than the reproducing process will be given. The signal drop due to loss of contact in the reproducing process can be calculated by the separation loss formula of 54.6 db loss per wave length of separation. With a 1 mil recorded wave length, a separation of 1/10 mil would cause a 5 db drop in signal. Obviously it would not take much dirt to cause severe drop-outs during the reproducing process. On particularly bad or dirty tapes, the drop-outs due to the reproduction process can become very large and numerous. However, under some conditions this same loss of contact distance of 1/10 mil can cause more than a 5 db drop in signal during the recording process. This may be explained by the use of Fig. 10 which shows the range of the output vs the bias curves for all the tapes used in obtaining the data for this study. In order to improve high-frequency response, most recording systems are set to use the lowest bias possible, perhaps in the region of A. Below this point the output decreases very rapidly for less bias. Limited measurements taken by the author have indicated that the effective bias in the tape itself is quite sensitive to the loss of contact between tape and head. A 1/10 mil separation can easily reduce the effective bias to about 50 per cent or even 25 per cent of its original value. At 50 per cent, the output could be down over 15 db while at 25 per cent it could be down by 30 db. Obviously, if drop-out effects are to be reduced in the recording process, the bias used should be higher on the output-bias curve.

It has long been assumed that the biased recording process might result in more drop-outs than in the saturated unbiased recording process as is used in FM carrier and digital applications. The previous paragraph points out how the changing effective bias level would affect the amplitude of the recorded signal. It would seem that if no bias is used, then this critical condition would not exist. However, actual tests of unbiased saturation systems indicate that the drop-out count is nearly the same as for a well biased system. This would seem logical since the curve of output vs recording signal level (unbiased) is somewhat like Fig. 11. The curves rise sharply just before saturation; therefore, if a loss of contact occurs while a normal saturating signal is being applied, the actual flux reaching the tape can easily drop low enough to reduce the output considerably.

Qualitative Rankings. One further example of the way in which the drop-out data can be and is being used is in the relative ranking of tapes according to analog and digital capabilities. Such a ranking is only a very crude first step in trying to arrive at some kind of classification system, and was used only because of a special requirement.

This particular rating was done to supply data to the National Bureau of Standards to help determine the correlation between the drop-out measurement technique and a Bureau of Standards chemical technique of testing the quality of tape.

Tapes were grouped into 5 classifications, A (Excellent), B (Good), C (Average), D (Poor), and F (Very Poor). The table of rating that resulted from

such a treatment is shown in Table 3. At least a definite idea can be obtained concerning the relative merits of tapes for analog and digital applications, though these ratings are not proposed as, or can they be considered as rigidly correct. The kinds of figures encountered and the rating scale used to arrive at the grades listed are shown in Table 4. It must be repeated that these figures should not be used as an absolute guide to classify tapes.

Conclusions

As a result of over two years of testing tapes and machines, measuring drop-outs with the equipment described, using the chart recorder, constantly monitoring the output signal on an oscilloscope, and noting the oxide build up on heads, rollers, and guides, the following conclusions have been reached:

1. Reel effects on these short-term drop-outs is relatively minor. Warped and eccentric reels would usually affect several inches of tape so the variations in output signal would be relatively long. However, the shorter drop-outs can be affected by the way a tape winds on itself on a reel. If an edge of the tape protrudes from the stack, it may become scalloped and thus make poor contact when pulled across a head. It is recommended that tape never be left in a high speed wind condition.

2. Environmental dirt in a normal laboratory atmosphere is relatively minor. Tests made with tapes that had no oxide to shed, showed remarkable consistency run after run, and even drop-out to drop-out.

3. Difference in drop-out counts on different machines, excluding any obvious physical damage to the tape, are determined by the manner in which the transport controls the tape as it is drawn over the heads. Transports with tight loops are generally better than those with open loops, though the wear is usually more severe. It would seem reasonable to assume that the tighter the tape is held against the head, the less effect a tape imperfection or physical deformity would have on the signal output, while at the same time more wear could be expected. Machines do not generate drop-outs.[1] They do change the effective size of any given drop-out.

4. Self dirt in the form of oxide flaking and build up is certainly one of the main causes of increased drop-out counts with use. If it is assumed that a new tape has a clean surface (and this assumption is by no means always valid) the first run of a tape should be a good indication of the imperfections of the surface of the tape. As tapes are run repeatedly, flakes of oxide coating wear off or flake off and can become attached to the heads, the rollers, and pressure idlers, or to the tape itself. Photographs of oxide particles building up on tape surfaces have been published by many researchers. Such self-dirt does play an important part in the drop-out history of a tape. If the machine surfaces are kept thoroughly clean, the drop-out count tends to remain more consistent. However, even machine cleanings after every 1000 ft pass of tape fail to prevent the oxide from building up on the tape itself. One possible solution to this problem is the continuous tape cleaning of the tape as it is passing through the machine as some machines now are equipped to do. However, it has yet to be proved by quantitative data just what is the most effective cleaning technique. Various methods of cleaning are being investigated using the drop-out measuring equipment. While the information is still not complete, results so far show that ultrasonic cleaning techniques improves very dirty tape, especially where large drop-outs are numerous, but that small drop-outs may even increase. Preliminary tests of a vacuum cleaning unit show very little effect on drop-out count. It would be expected that loose particles would be removed, but it has been shown by others that oxide particles are usually fused onto the surface of the tape, apparently by the pressure and heat to which a tape is exposed while going through a tape drive system. Knife edge cleaning systems have not been tested as yet. It would seem that such a

[1] If machine guides and rollers are not properly aligned with the reels, drop-outs on edge tracks can increase due to physical action on the tape edges.

method could remove some fused flakes but the danger of physically injuring other areas of the tape must be considered.

Scope of Present and Future Efforts

In addition to a study of various cleaning methods, wide tapes are being investigated to determine channel to channel correlation and effects on various channels due to wear on various types of machines, and efforts will continue to simplify the amount of data required to properly evaluate any given recording system for a specific type of application. There has been considerable evidence that some recent tapes have been remarkably improved so it is expected that during the next few months, another complete testing will be made of all current 1962 production instrumentation and digital tapes.

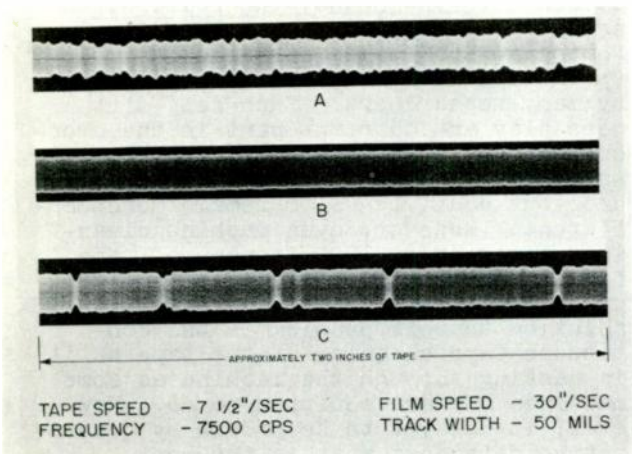


Fig. 1. Modulation envelopes of tape—past and present.

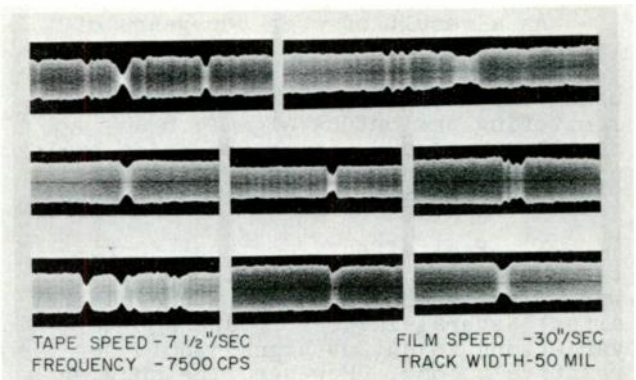


Fig. 2. Types of drop-outs.

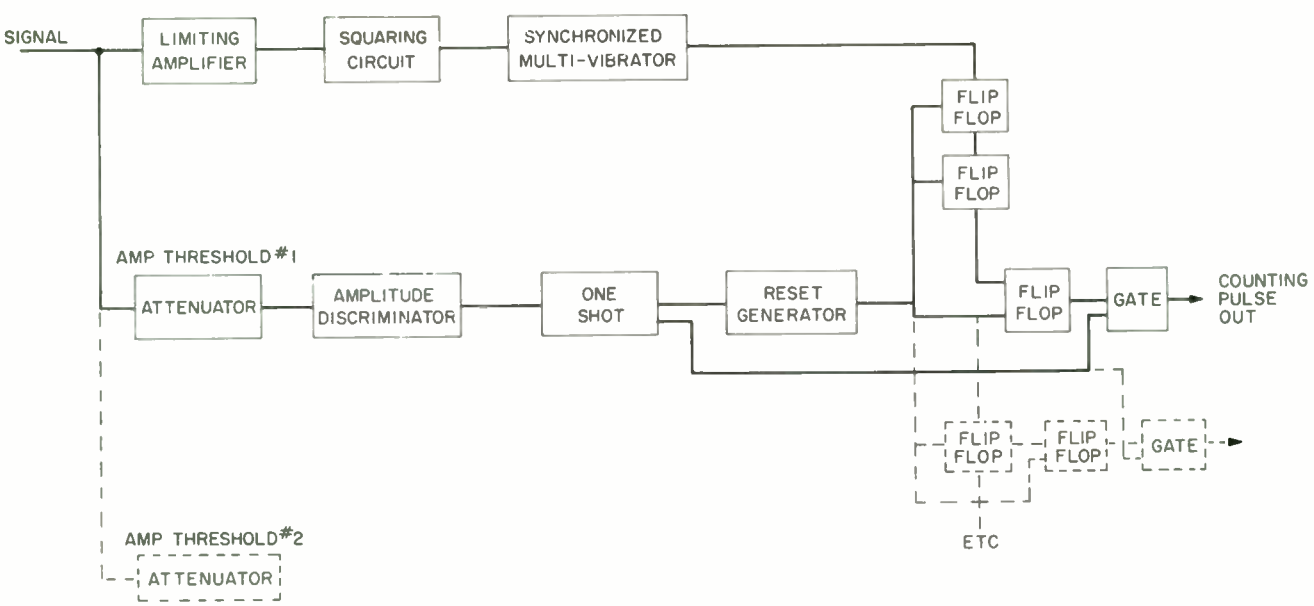


Fig. 3. Block diagram of basic drop-out measuring system.

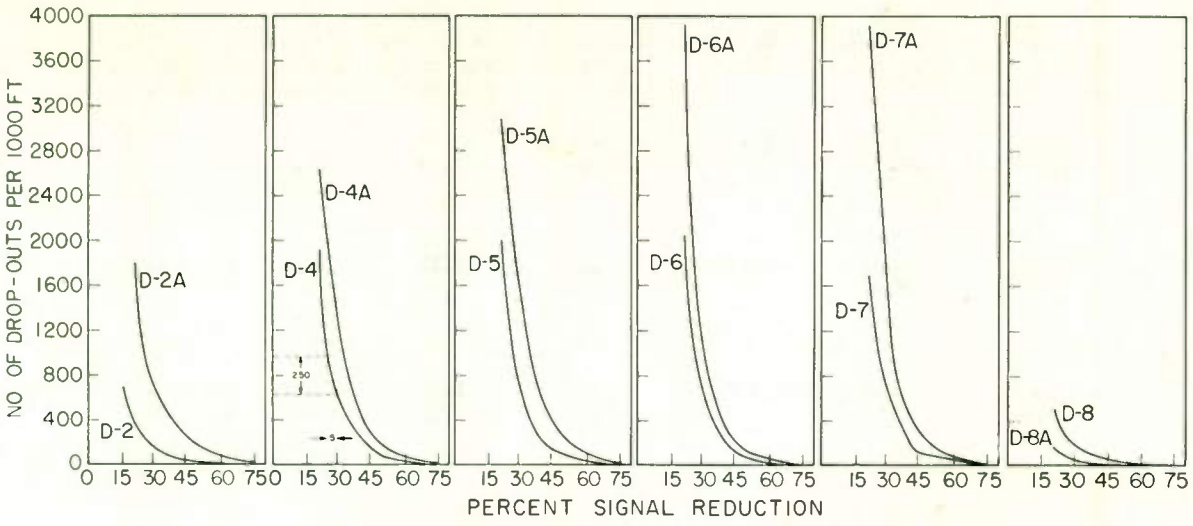


Fig. 4. Amplitude distribution of drop-outs.

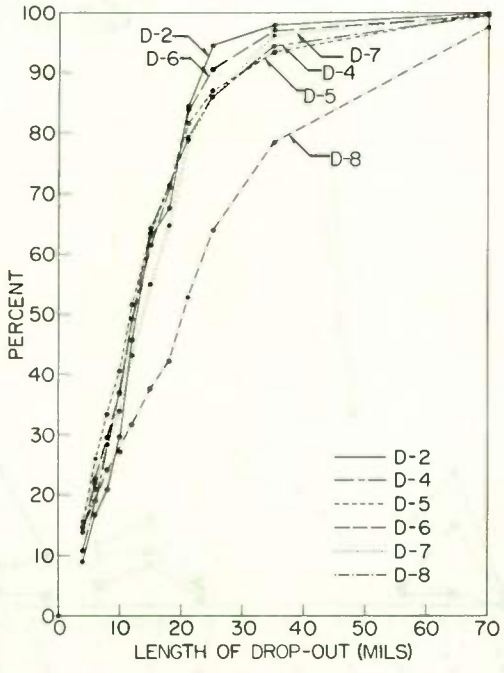


Fig. 5. Length distribution of drop-outs.

TABLE 1. Combined Drop-out Data for All Reels of Tape of the Same Type.

Amp. Thres. db	TOTAL NUMBER Mean and Std. Dev.	REEL STD. DEV. Mean and Std. Dev.	SLOPE Index	MINIMUM MAXIMUM Values	PER CENT of total no. Mean and Std. Dev.	PER CENT of total for each length shown Mean	Std.Dev.	Length Mils
3	153.60	32.45	22.90	57.00	84.66	5.75	1.92	36-560
	37.20	24.68		286.00	3.79	40.28	4.87	16-35
						30.00	3.17	9-15
						23.94	4.78	3-8
6	23.95	5.35	24.55	3.00	14.15	12.13	7.00	36-560
	8.60	2.34		41.00	3.22	56.78	8.06	16-35
						14.82	5.60	9-15
						16.23	5.26	3-8
12	1.85	.04	-7.78	.00	1.16	10.16	20.33	36-560
	1.20	.57		5.00	.72	49.16	35.43	16-35
						7.32	10.41	9-15
						13.33	20.97	3-8

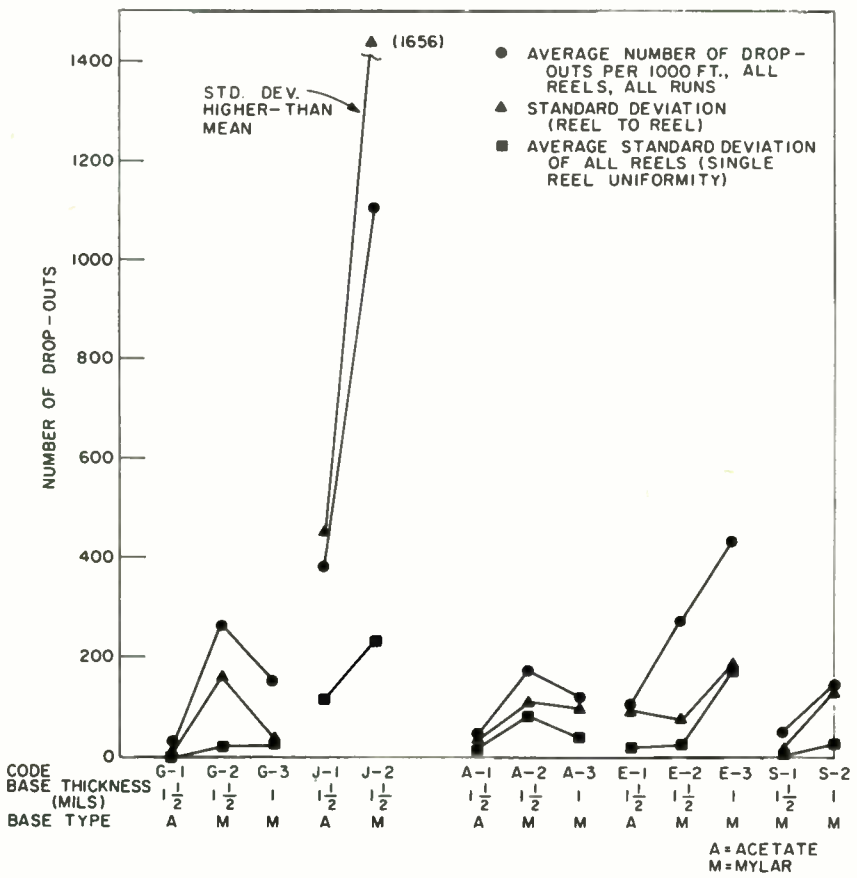


Fig. 6. Number of 3 db drop-outs and standard deviations for 5 reels of each type of tape—1959 tapes.

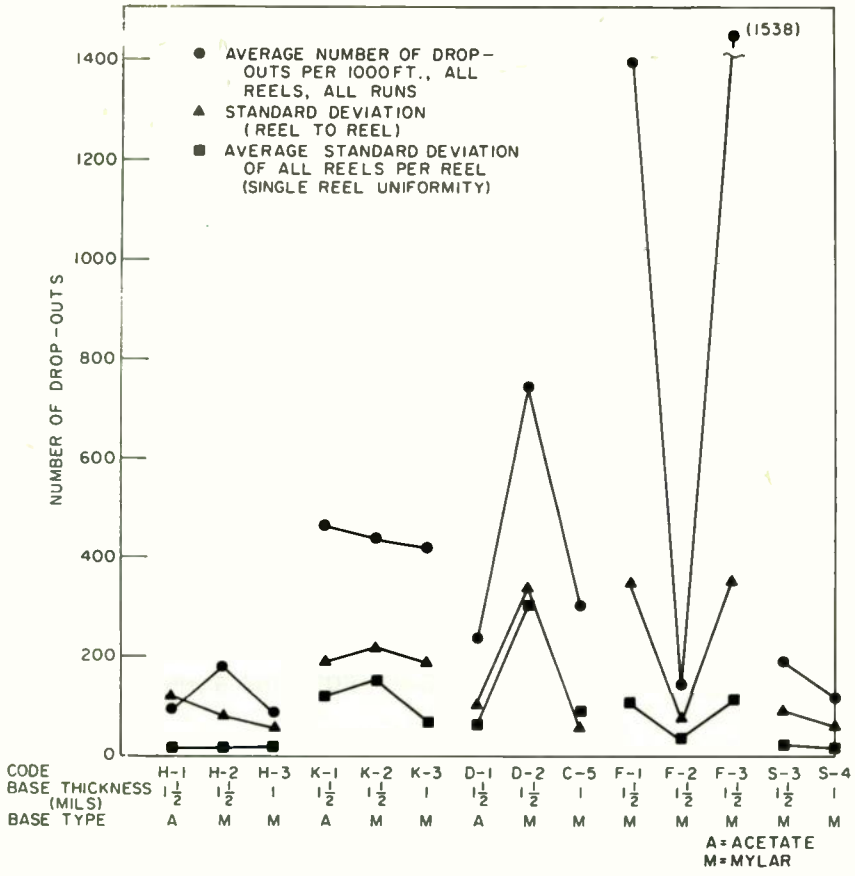


Fig. 7. Number of 3 db drop-outs and standard deviations for 5 reels of each type of tape—1960 tapes.

TABLE 2.

Repeatibility of Drop-Outs on Successive Runs (Percent* of Total Number)

No. of Runs the Same Drop-Out Occurs	Reproduce Each Run	Record and Reproduce Each Run
4	70	45
2 or 3	20	35
1	10	20

*Average percent from 4 tapes.

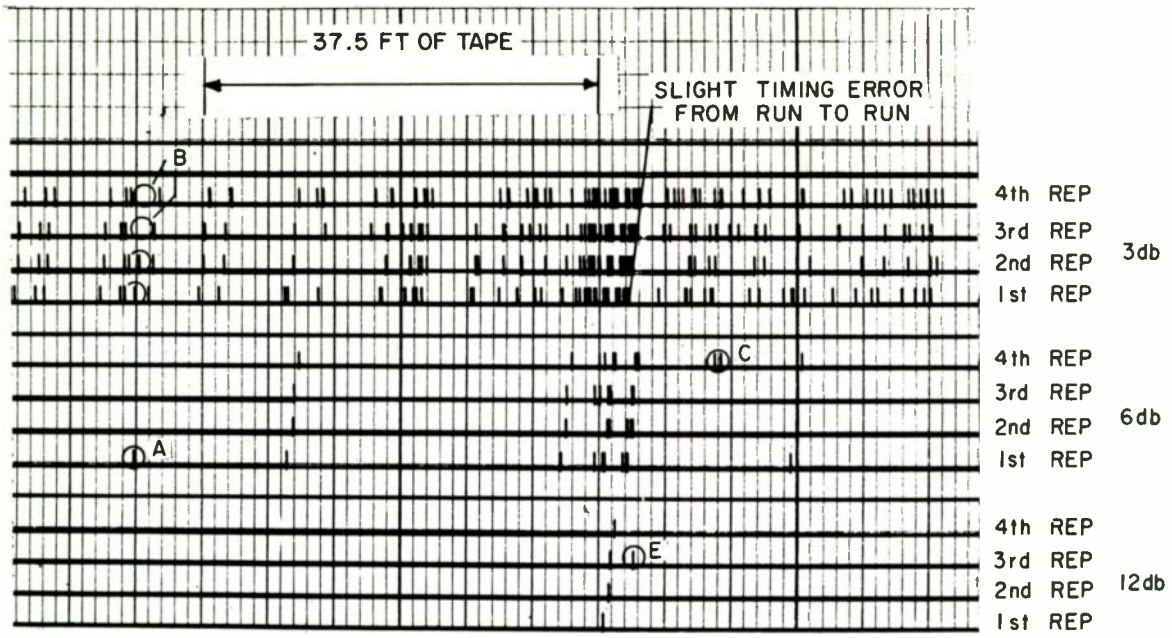


Fig. 8. Small section of chart. Sample recorded once and reproduced 4 times.

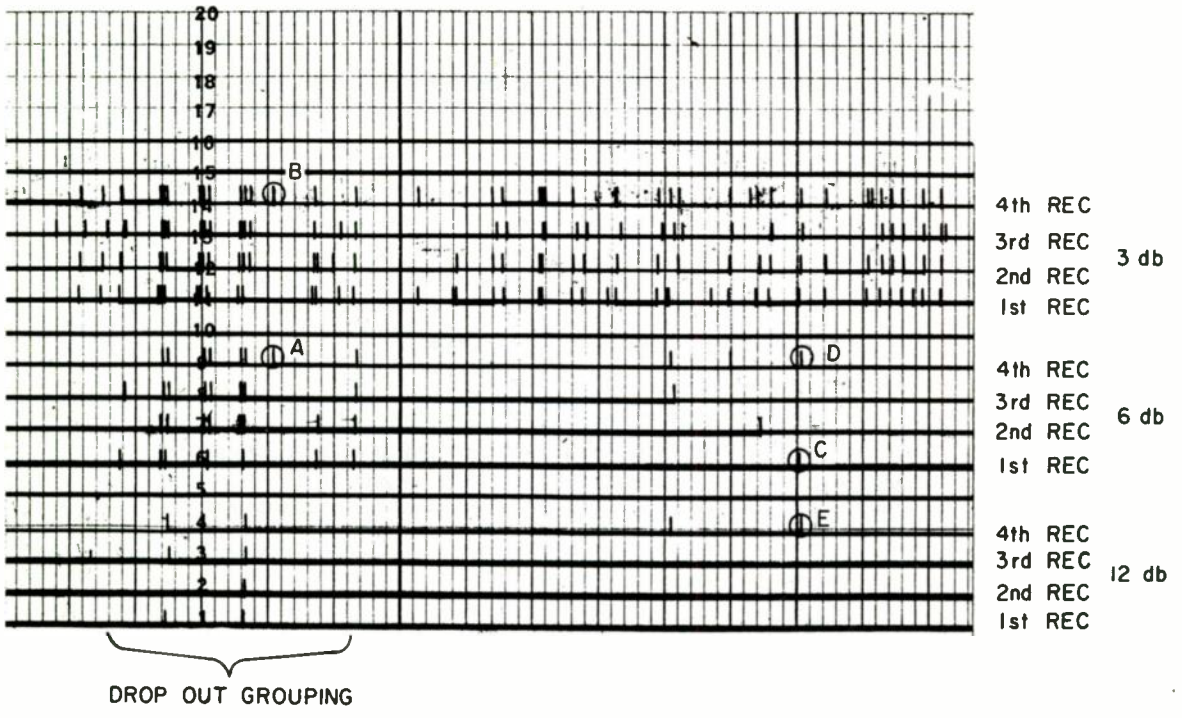


Fig. 9. Small section of chart. Sample recorded and reproduced each run for four runs.

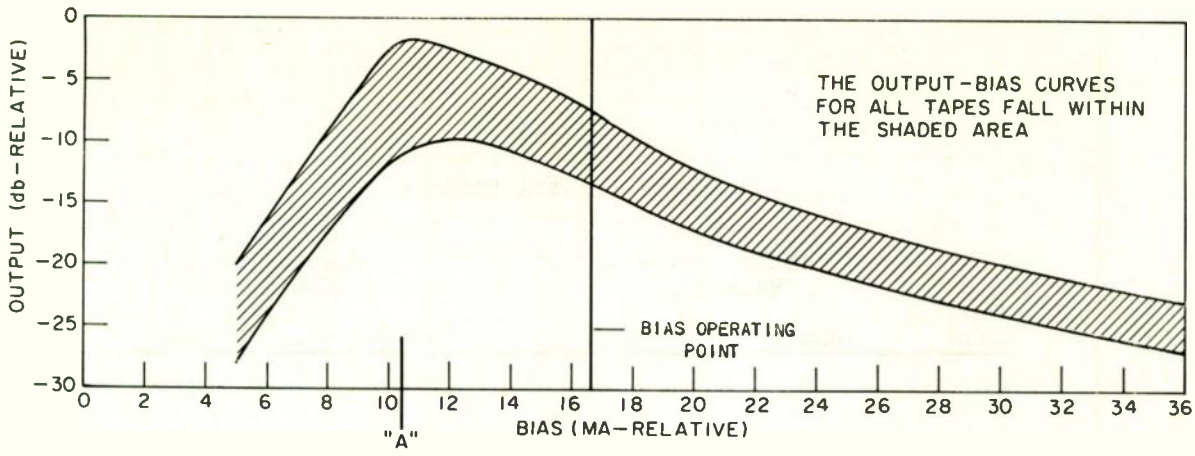


Fig. 10. Output-bias curve limits for all tapes used.

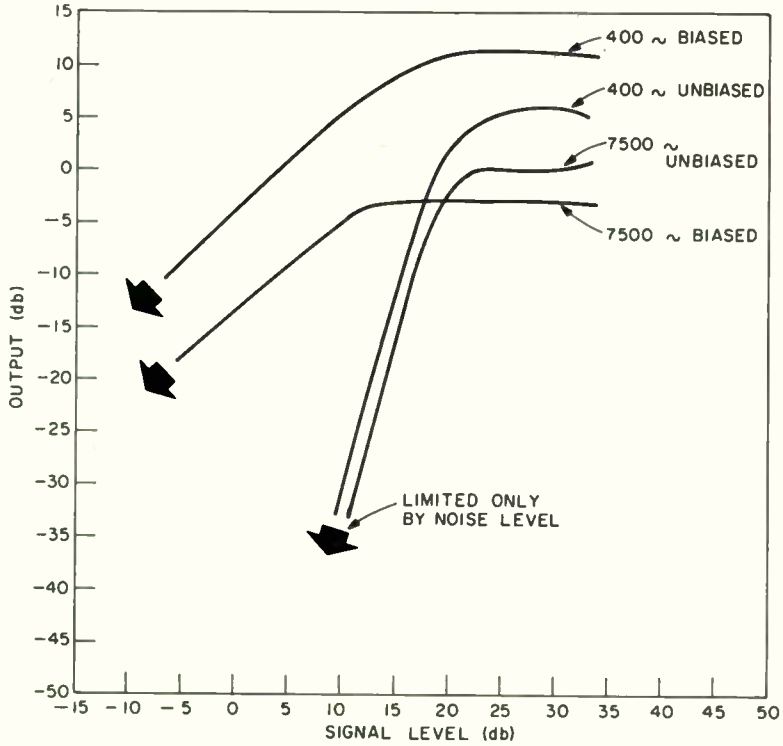


Fig. 11. Output vs input for biased and unbiased recording signals.

Table 3

Analog Digital Ratings

Tape Code	Initial Value Ratings		Wear Value Ratings	
	Analog	Digital	Analog	Digital
C-1	B	A	C	C
C-2	A	A	C	A
C-3	A	A	D	D
C-4	D	F	B	B
C-5	C	C	B	A
D-1	C	A	B	D
D-2	D	F	B	B
F-1	F	C	C	F
F-2	B	C	A	B
F-3	F	F	B	A
H-1	A	C	B	A
H-2	B	A	B	C
H-3	A	A	B	C
J-4	D	C	B	B
K-1	C	C	A	A
K-2	D	F	B	B
K-3	C	B	C	A
K-4	C	D	C	B
K-5	D	F	D	D
S-3	B	A	C	D
S-4	B	C	A	B

Table 4

Rating Scale for Analog - Digital Ratings

	Initial Values in No. per 1000 ft		Wear in per cent
Rating	3 db	12 db	3 db and 12 db
A	0-100	0-1/4	Any - value to +2
B	100-200	1/4-1	2 to 15
C	200-500	1-4	15 to 50
D	500-1000	5-10	50 to 150
F	Over 1000	Over 10	Over 150

SOME EXPERIMENTS WITH MAGNETIC PLAYBACK USING HALL EFFECT SENSITIVE ELEMENTS

Marvin Camras
Armour Research Foundation
of Illinois Institute of Technology

Summary. Flux sensitive heads using thin semiconductors have been developed for playback of magnetic recordings. The elements have a frequency response inherently flat from d-c to tens of megacycles, so that the main limitations are in the associated head structure. Factors entering into optimum design are discussed, including materials and configurations of semiconductor elements and cores. Experimental results for successful designs are given, and a number of applications are suggested.

INTRODUCTION

Pickup heads commonly used for playback of magnetic recordings are sensitive only to rate of change of flux picked up from the tape, and hence have an important basic characteristic: that as the frequency decreases the output falls off in proportion, becoming zero at zero frequency.

This results in the familiar constant-current response curve of Fig. 1A. To compensate for the drop at low frequencies, the amplifiers used with standard playback heads require a bass-boost amounting to 30 or 40 db at 30 cycles. As the frequency is decreased still further a point is reached where the head output is too low to be useful, and some other method must be found for sensing the recording on the tape.

One application which requires a very low frequency response is where data is recorded on tape at normal speeds and played back very slowly for analysis.[1] Another application is where control or reference pulses are recorded which must be picked up and counted at different and variable speeds even when the record slows down, stops, and starts again.

A number of approaches have been suggested for pickups that would be sensitive to magnetic flux rather than its rate-of-change.[2] These have an inherently flat response down to zero frequency as in Fig. 1B. One class of flux sensitive detectors is based on the idea of interrupting or modulating the flux from the record at a fairly high rate, thus giving a rate-of-change of flux which can be sensed with a coil. Motor driven vanes and vibrating reeds in the head core were proposed, but the result can be achieved much more simply by the magnetic modulator principle, where the signal flux is modulated by means of auxiliary windings on a saturable portion of the head.

Another class of detectors utilizes magneto-electric effects other than electromagnetic induction, including magnetoresistance, hall-effect, the deflection of electron beams[3], and the (slight) sensitivity of transistors to a magnetic field.

Still other miscellaneous magnetic effects are available, such as rotation of polarized light beams[4], magneto-mechanical forces, and changes in incremental permeability.

Of all the above, the magnetic modulator principle has been probably the most successful. However it does have some limitations: a source of high frequency excitation is required; there is background noise due to the Barkhausen effect; an upper limit to the frequency response is inherent in the ferromagnetic core structure; and windings are still required as in conventional heads.

Hall-effect sensitive elements do not have the above limitations, (although they have a few of their own, as will become apparent). The possibility of low manufacturing cost and the inherently flat playback response makes it conceivable that such heads might replace conventional heads; which up to the present have not had a competitor except for special and expensive instrumentation uses. Accordingly, hall-element heads were investigated both analytically and experimentally, to determine whether an advantageous design was possible.

THE HALL EFFECT

The Hall-effect is named after E. H. Hall, who discovered in 1879 that he could skew the equipotential lines in an electrical conductor by applying a magnetic field.

The Hall-effect is shown in Fig. 2. A conductor of width W and thickness T carries a current I. With no magnetic field, if a voltmeter V is connected to opposite edges on the same equipotential line it reads zero. When a magnetic field H is applied at right angles to both the current and to the voltage axes, the voltmeter gives a steady deflection. The magnitude of this hall-voltage is

$$V_H = \frac{R\,I\,H}{T}$$

where R is the hall-coefficient characteristic of the material.

The above formula satisfies our intuitive expectations that the generated voltage should be proportional to field strength (H), to density of current (I/WT), and to separation (W) between the voltage pickup points.

Ordinary conductive materials have a Hall-effect which is too small to be of value as a sensor for magnetic fields. But certain semiconductors have hall-coefficients which are orders of magnitude higher, and these have made it practical to sense fields of magnetic tape recordings.

PLAYBACK HEADS

A Hall-effect playback head may have the same core structure as a conventional head, but instead of the winding, a semiconductor element is placed where magnetic flux from the tape can act on it. Figure 3 shows a hall-element in the back gap of a pair of C-shaped cores, with the front gap contacting the tape recording. There are four terminals. Two of these are energized by a steady dc current of perhaps 50 to 500 milliamperes. At the other two terminals, an output voltage appears which depends on magnetic flux from the tape that is picked up by the core. This is true even if the tape is not moving, in which case the output is d-c of positive or of negative polarity depending on the direction of core flux. Figure 4 is a photograph of an experimental head of this kind.

But a conventional core is not necessary, and in fact a coreless (as well as coil-less) head is possible as in Fig. 5A. Here a miniature semiconductor wafer is mounted so that one of the thin edges bears against the tape. The field of the tape acts directly on the element. In the absence of coils and cores the high frequency response of this head is limited only by the Hall-effect; and the Hall-effect operates at frequencies as high as 10,000 megacycles (10 Gigacycles)[5]. Ferromagnetic effects such as non-linearity, hysteresis, and Barkhausen noise are absent.

While the elegance of a coreless design may have special appeal to mathematicians and to high-fidelity enthusiasts, omission of the core is detrimental to resolving power and to sensitivity of the head. If we do not need the full 10,000 megacycle response (and we probably won't until the mechanical engineers give us faster tape-drives) then we can back up the element with ferrite pieces as in Fig. 5B and still go to about 100 megacycles. Or we can use thin metal cores if the upper frequency limit does not exceed a few megacycles. For convenience, the above heads are designated as "front gap" heads.

Heads as in Fig. 5 where the element is next to the tape, have the problem of how to connect to the edge of the semiconductor on which the tape rides. Several methods of making this connection have been proposed. However, the construction and operation of the head is greatly facilitated if the connection is eliminated altogether. Three ways that we have tested are shown in Fig. 6. In Fig. 6A a phantom equipotential point is established by connecting one of the output leads to a tapped resistor in place of the missing side-terminal. This arrangement loses about half the output voltage. Figure 6B is five terminal element where the exciting current flows in opposite directions from the center to each end; hence the output terminals b, d can be at the same quiescent potential, but acquire opposite potentials when a magnetic field is present. Best results from the standpoints of output and low noise were obtained with a K shaped element as in Fig. 6C.

FACTORS FOR OPTIMUM DESIGN

For playback heads we are interested in linearity, sensitivity, and a low noise-level. The linearity is practically perfect up to the highest fields likely to be encountered in magnetic recording, as is shown by the experimental curve of Fig. 7 which is typical of hall-elements.

The sensitivity is determined by geometry of the head and element; and by the electromagnetic properties of the element. Referring again to Fig. 2 and to the equation for hall-voltage:

$$V_H = \frac{R\,I\,H}{T}$$

it would appear that we should seek a material with the highest possible hall-coefficient (R). Experience indicates that this is not sufficient, because resistivity (ρ) of the material is also important in two ways: (1) lower resistivity gives less heating and allows the exciting current (I) to be increased; and (2) lower resistivity means that the output voltage can be delivered into a lower impedance load. A good index of electromagnetic merit is the mobility, which is the hall-coefficient divided by the resistivity.

$$\text{Mobility} = \mu = \frac{R}{\rho}$$

Germanium has a mobility of approximately 2000, while the mobility of indium antimonide is approximately 30,000. Experience with heads using these semiconductors confirmed that indium antimonide was superior.

The exciting current (I) will normally be set at the highest safe value, since the output voltage goes up in proportion. A major limitation is the heat generated by ohmic losses in the exciting circuit. Semiconductor elements are

temperature sensitive, and the output level will drift as the head warms up. The exciting current is therefore determined by temperature stability as well as by possibility of burnout.

The field (H) results from whatever tape flux can be channeled through the hall-element. A quarter inch tape at saturation can supply about 1/2 maxwell; but at normal recorded wavelengths, intensities and track widths a realistic estimate would be 1% to 10% of this amount. Leakage and core losses reduce this still further, so that the element operates at very low levels. An efficient magnetic design is more important in Hall-effect heads than in ordinary types because the low output voltages which are a fraction of a millivolt do not allow a margin of safety. Leakage and coupling losses should be minimized.

Thickness (T) appears in the denominator of the Hall-voltage equation, and hence should be minimized. Reducing the thickness of the hall-wafer also improves the magnetic efficiency by decreasing the magnetic circuit reluctance. These benefits are accompanied by a smaller cross section in the element with higher ohmic resistance, so that the exciting current (I) must be reduced. Since heating is varied as I^2, the current needs to be reduced only in proportion to the square root of the higher ohmic resistance, and there is a net advantage in thin elements.

EXPERIMENTAL RESULTS

In earlier experiments, we made the Hall-elements from germanium. Polycrystalline and single-crystal bulk materials were tried, as well as evaporated films. The single crystal variety turned out to be the most sensitive. A number of heads were built with single crystal germanium, some with elements in the back-gap as in Fig. 3, and others with front elements shaped as in Fig. 6. All of these designs performed satisfactorily. Of the front-gap types, the best results were obtained with the K element of Fig. 6C.

A characteristic unequalized frequency response curve of a Hall-effect head is shown in Fig. 8. Here the record was made at constant recording current throughout the audio range. Low frequency output remains constant, and actually goes all the way down to zero frequency. Those who are familiar with constant current curves of ordinary heads (Fig. 1A) will be disappointed at the apparent "high-end" response of Fig. 8 above 1000 cycles, until we point out that it is really better than Fig. 1A if we correct for the 6 db per octave rate-of-change rise in Fig. 1A.

It is clear that low frequency equalization may be omitted for a hall-head, but that high frequency equalization is required to bring its performance to the same level as a conventional head having the same core and gap.

With the germanium elements a noise was generated by passage of the exciting current. Measurements showed that the noise increased in direct proportion to the current, so that when enough excitation was used to bring this noise well above the background noise of the amplifiers, there was no advantage in further increase of excitation since the signal and noise both rose together.

Other semiconductors were also tried in our Hall-effect heads. Indium-antimonide was the best of these by a wide margin, with respect to high sensitivity and low noise. Using this material, we obtained a signal-to-noise ratio of more than 50 db over audio bandwidths. For audio work, the variation in sensitivity with temperature change was not noticeable. In instrumentation, where drift is more important, alternative semiconductors may be used at a sacrifice of output for the sake of stability.

Dc excitation of the hall-element is by far the simplest, but when a system must respond all the way down to zero frequency, then we require stable d-c amplifiers after the head output, and these are sometimes troublesome. Alternatively, we can energize the hall-element with a-c instead of d-c, in which case the output will be an a-c signal of the excitation frequency whose amplitude depends on the amplitude of the signal picked up by the head, and whose phase depends on the polarity of the signal. Carrier techniques are then applied; including narrow-band amplification, and demodulation to give the analog of the signal. A-c systems were investigated and were found practical. The chief limitations were the additional circuitry required, and the stability against drift from a balanced condition.

APPLICATIONS

Hall-effect heads are the simplest form of detector for pickup of low frequency information from slowly moving tapes. This is advantageous in data analysis, control systems, and instrumentation.

Output of the head is an analog of the recorded magnetic flux. Thus complex waveforms may be reproduced and displayed without phase distortion.

Static sensitivity to very small magnetized areas, in the order of the gap dimensions, makes the hall-head a useful tool for study of the magnetic recording process itself.

High frequency limitations and non-linearities of ferromagnetic core and coil construction may be avoided by certain hall-head designs.

Hall-effect heads are also useful for general purposes as for example high fidelity audio reproduction; but their application in this field depends on economies of production, and relative circuit advantages compared to present day heads.

ACKNOWLEDGEMENTS

The author is grateful to Dr. J. J. Brophy for his suggestions and for making various semiconductor materials available to this project; to Mr. Carl Christensen who fabricated the elements and heads; and to Mr. Philip Padva who tested the heads.

REFERENCES

1. M. Camras, "Tape Recording Applications", IRE Transactions on Audio, V. AU-3, p. 178; Nov.-Dec., 1955.

2. O. Kornei, "Survey of Flux-Responsive Magnetic Recording Heads", Journal of the Acoustical Society of America, V. 27, p. 575; May 1955.

3. A. M. Skellett, L. E. Leveridge, and J. W. Gratian, "Electron Beam Head for Magnetic Playback", Electronics, V. 26, p. 168; Oct. 1953.

4. A. W. Friend, "Magneto-optic Transducers" RCA Review, V. 11, p. 482; Dec. 1950.

5. D. P. Kanellakos, R. P. Schuck, and A. C. Todd, "Hall Effect Wattmeters", IRE Transactions on Audio, V. AU-9, p. 8; Jan.-Feb., 1961.

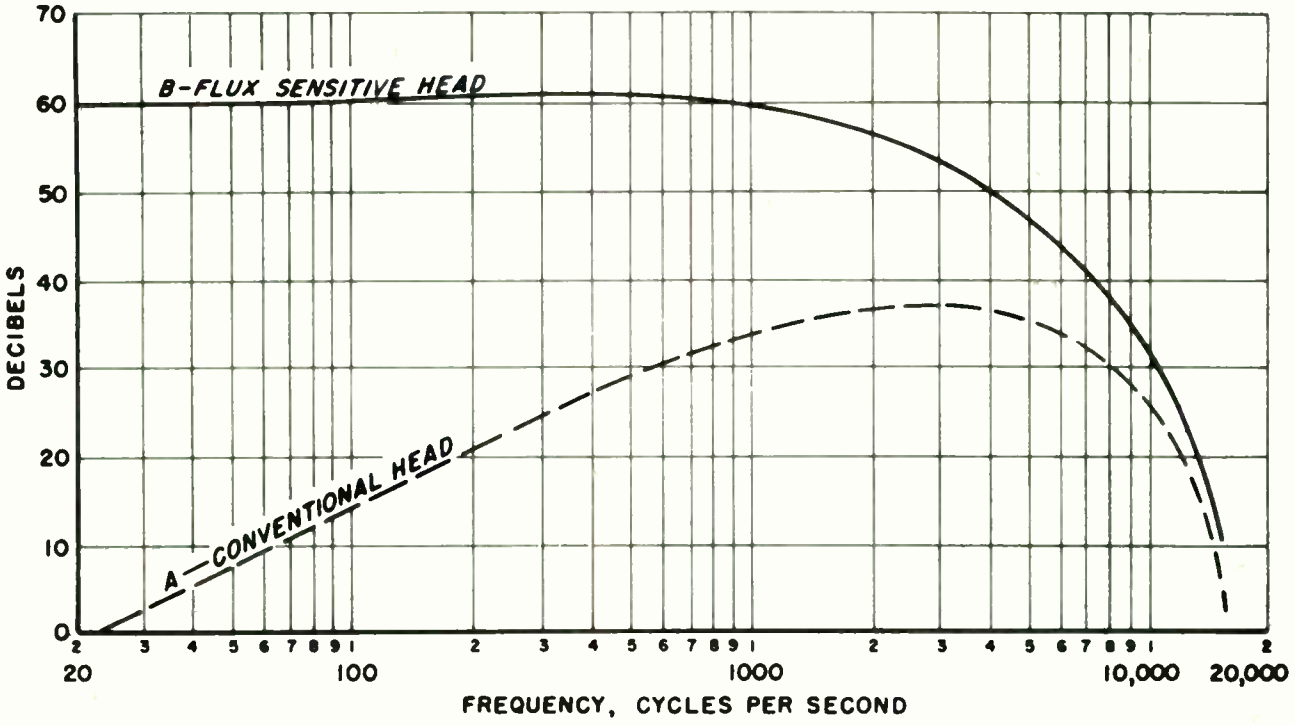
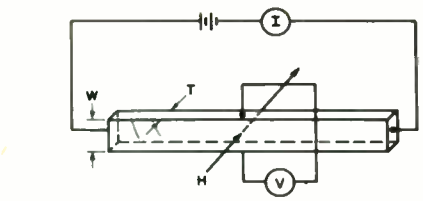


Fig. 1. Unequalized frequency response of two types of heads when playing back the same constant current recording.



HALL VOLTAGE, $V_H = \frac{RIH}{T}$

I = CURRENT
H = MAGNETIC FIELD
T = THICKNESS
R = HALL COEFFICIENT = $\frac{1}{ne}$ (n = NO. OF CARRIERS, e = CHARGE PER CARRIER, u = MOBILITY OF CARRIERS)
ρ = RESISTIVITY = $\frac{1}{neu}$
$\frac{R}{\rho} = u$
u = 2000 FOR GERMANIUM
u = 30000 FOR InSb

Fig. 2. The hall effect.

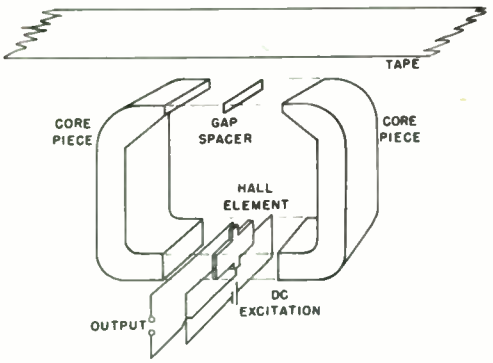


Fig. 3. Elements of a back-gap hall-effect head.

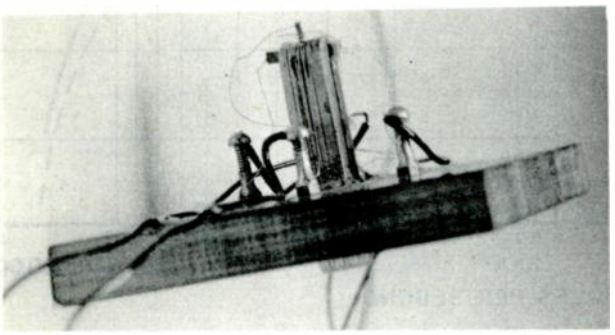


Fig. 4. Experimental head.

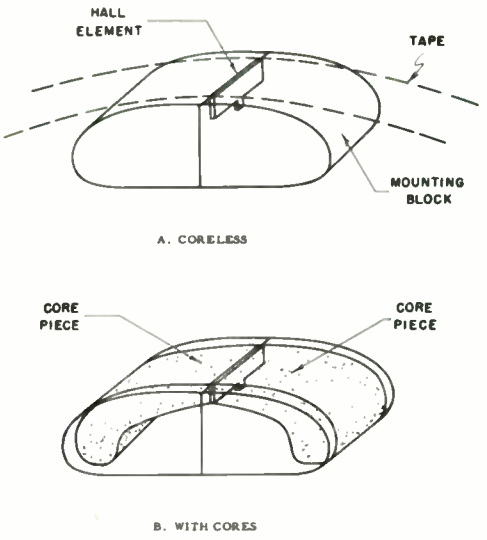


Fig. 5. Front gap hall-effect heads.

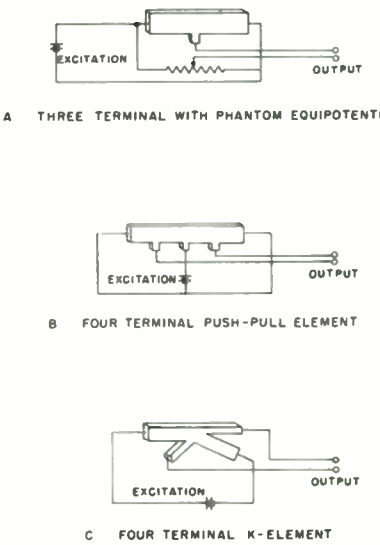


Fig. 6. Circuits and elements for front-gap heads.

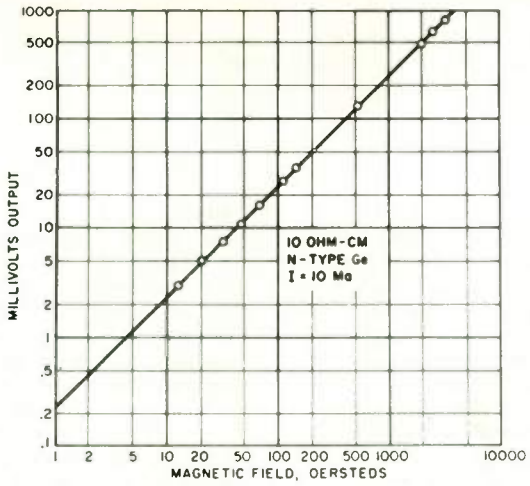


Fig. 7. Output vs. field for germanium hall element.

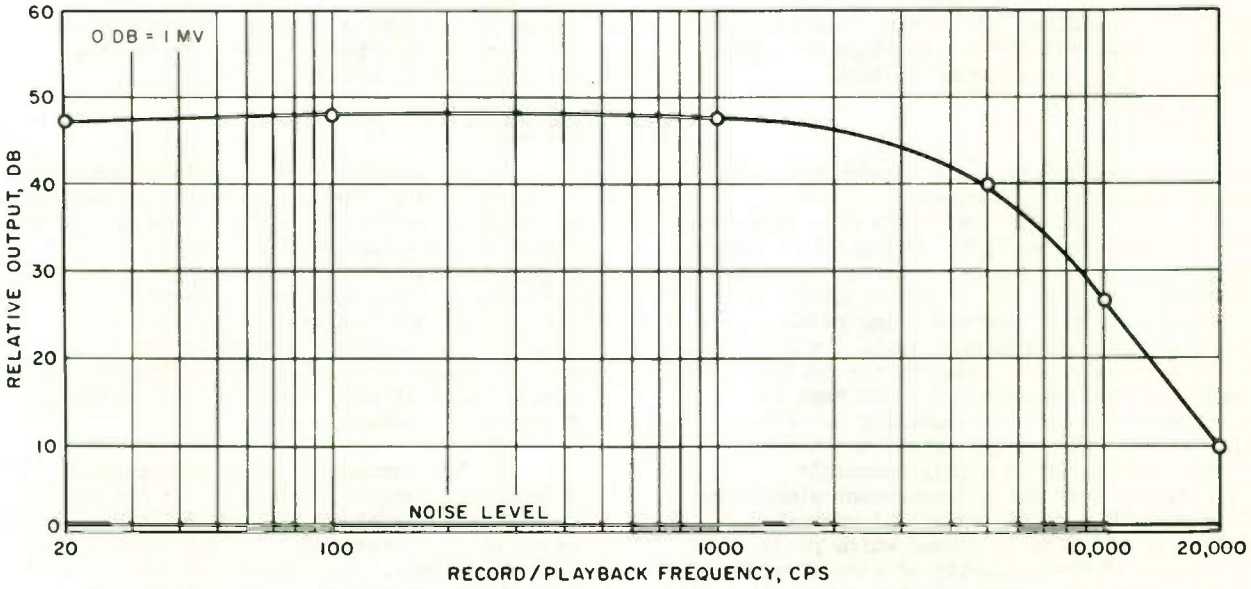


Fig. 8. Hall effect head, playback output.

ABSOLUTE MEASUREMENTS OF MAGNETIC SURFACE INDUCTION

Frank A. Comerci
CBS Laboratories
Stamford, Connecticut

Summary

Two methods have been used for obtaining an absolute measurement of normal surface induction of a medium wavelength signal recorded on magnetic tape. One employs a nonmagnetic loop and an electrical procedure for determining its active rectangular cross-sectional dimensions. The normal surface induction is computed from these dimensions and the output voltage induced in the loop as the tape is transported at a prescribed speed across the active portion of the loop. The other is based on recording a dc signal of current level equal to the rms value of the ac recording current which produces the equivalent recorded signal as that for which the measurement is desired. A calibrated search coil is then used to measure the total remanent flux in a bundle of many layers of the dc recorded tape. The surface induction is then calculated for medium wavelength signals using existing formulae.

Measurements of surface induction obtained using the two methods provide equivalent results. The surface induction could be measured for a wide range of wavelengths using the nonmagnetic loop.

Introduction

Several years ago a nonmagnetic loop was shown[1] to be an adequate tool for measuring the surface induction of a relatively long recorded wavelength. Using this measurement, the surface induction for other wavelengths could then be determined using the "short gap method" for measuring relative levels. A subsequent publication[2] stated that the nonmagnetic loop measures the surface induction for the case where the tape is suspended in a unit permeability material, such as air, whereas in actual use the tape is in contact with a highly permeable reproducing head and a measurement simulating this condition is preferred. A measuring procedure was then described which provides a measure of the intensity of magnetization within the tape coating in terms of millimaxwells per millimeter of track width.

Details of the two methods have been adequately described in the literature. Actually, in either method of measurement, the magnetized tape is surrounded by air or a medium of unit permeability. When using the nonmagnetic loop, the field strength existing at the surface of the tape is measured directly and the intensity of magnetization within the tape coating can be calculated if the effective tape dimensions are known. The flux lines or millimaxwells through the tape cross section could be calculated directly. For the search coil method, the flux lines through the tape cross section are measured and the resulting surface induction for a relatively long recorded wavelength could be easily calculated. Any difference in measurements performed by the two methods excluding experimental error would then reflect on the validity of the formulae used for calculation.

A short experiment was conducted to investigate possible differences in the measurements. The results were essentially the same. However, an analysis of the data revealed a graphical technique which, under the conditions employed, was able to separate the magnetic losses.

Procedure

Measurements of Normal Surface Induction were obtained by the nonmagnetic loop and the search coil techniques for a single sample of magnetic recording tape at several bias and signal levels. The results were then compared. The data was then examined to determine if the losses could be segregated. The nonmagnetic loop results were then used to obtain measurements of surface induction for a wide range of wavelengths and to calibrate a magnetic reproducing head.

For the nonmagnetic loop measurements, a loop was constructed as shown in Figure 1. Two rectangular blocks of polishable insulating material were fabricated to 5/16 x 5/8 x 7/8 inch dimensions. The larger surfaces intended to mate when the two blocks were clamped together were ground and polished optically flat. Two grooves were then milled close to the long edge of the polished surface of one block of sufficient size to

accept copper conductors which formed the rear portion of the loop. A piece of platinum foil 1 micron thick was then cemented over the polished surface of the milled block bridging the front ends of the two conductors. The foil was soldered to the two conductors at the junction points. The rear portion of the foil was then ground away to a distance 10 mils from the front edge of the block. In performing this operation a deeper cut was made near the edges of the block to form a reference mark locating the rear (covered) edge of the foil. The two blocks were then bolted together with the polished surfaces mating. The front surface of the structure was then ground to a 1/2 inch radius parallel with the anticipated foil edge to a point where the dimension of the platinum foil, front to back ("d" dimension), as determined by the distance from the front surface to the reference marks, measured 5 mils. The front surface was then hand lapped to a final finish. The cross-sectional dimensions of the remaining platinum material which formed the active part of the loop were measured by microscope to be approximately 4.5 mils deep ("d" dimension) and 1 micron (39 microinches) broad. ("b" dimension)

A sample of 150 mil wide magnetic recording tape known to have an exceptionally smooth flat surface and a coating thickness of 0.35 mils, was selected for the measurements in an effort to avoid excessive head to tape spacing and thickness loss. Full track recordings were made on this tape at several wavelength increments from 93.5 to 0.187 mils. The recordings were made with constant signal current and a bias level (140 kc) which gave peak signal output for a wavelength of 0.935 mils. The tape speed used was 1.87 ips. This low speed was selected to avoid using unnecessarily high frequencies with corresponding electrical and core losses. The signal current used was that which resulted in 5% distortion in the non-equalized playback output for a 0.935 mil recorded wavelength. This level is hereafter referred to as the "maximum recording level". The bias current will be referred to as "optimum bias". The record head was known to provide equal magnetizing fields for constant current input for the signal frequency range employed. It had a gap length of 2 microns. The optimum bias was judged to record through only 0.25 mils of the tape thickness on the basis of the ratio of the increase in output which could be achieved when the bias was increased for a long wavelength signal. This underbiased condition was used to meet restrictions of the search coil technique.

A playback head with a 1 micron gap was used to measure the output from the recorded tape. This head was of a design which had proved to be very stable over a use period of more than one year and could serve as a standard head after being calibrated.

The recorded tape was then operated on a 15 ips tape transport on which the non-magnetic loop was mounted. The output voltage of the loop was measured for each recorded wavelength. In order to measure the extremely small voltages from the loop a high step-up-ratio transformer and appropriate band pass filters were required in the measuring circuits. Calibration of the measuring circuit was achieved by introducing known voltages across a one ohm resistor in series with the loop and transformer primary. Even with this arrangement the low levels at longest and shortest wavelengths could not be measured because of the low signal to noise ratio.

The effective depth of the active loop cross section was determined by methods described in the literature.[1] A plot of the loop output voltage vs recorded wavelength was made and the wavelength noted for an output which in the longer wavelength direction was 3 db less than that over the flat portion of the curve. The depth dimension was then calculated from the simple expression

$$d = 0.19\ \lambda_{3db}$$

where d = depth dimension (mils)

λ_{3db} = wavelength for 3 db long wavelength loss (mils)

The normal surface induction at a wavelength of 4.7 mils for which short wavelength losses could be neglected was then determined using the formula

$$By = \frac{e_{max} \cdot 10^8}{v\ w \left[\frac{1 - e^{\frac{-2\pi d}{\lambda}}}{\frac{2\pi d}{\lambda}}\right]}$$

where By = normal peak surface induction (gauss)

e_{max} = peak output of loop (volts)

v = tape velocity (cm/sec)

w = width of recording (cm)

d = depth dimension of loop (mils)

λ = wavelength (mils)

For the search coil measurement the same tape sample used for the nonmagnetic loop measurement was recorded using the same recording equipment but a dc record current equal to the peak value of the constant current maximum recording level used previously. Recordings were made for optimum bias conditions at four dc current levels and at the maximum dc current level for four bias levels. The output of the playback heads was also measured when corresponding rms values of ac current to produce a 4.7 mil wavelength recording were used instead of dc current. These additional measurements were made to observe any differences between the normal head playback results and the search coil results for different shaped recorded wavefronts or direction of magnetization.

Each of the dc recorded sections representing the different bias and dc levels were cut into 30 five inch lengths which were placed one over the other to form a bundle of 30 lengths. Care was exercized to be certain that the magnetizations of each length in the bundle were in the same direction. The bundle of samples were inserted into a search coil similar to that described in the literature.[3] The search coil was connected to an electronic integrating amplifier and meter. The bundle was then quickly withdrawn from the coil and the remanent flux ϕ_t in the bundle was read on the calibrated meter. The remanent flux ϕ for a single layer was then equal to $\phi_t/30$.

The surface induction at a wavelength of 4.7 mils was then calculated using the formulae[3]

$$J_x = \frac{\phi}{4\pi wC} \quad \text{and} \quad B_y = \frac{4\pi^2 C J_x}{\lambda}$$

from which $B_y = \frac{\pi\phi}{w\lambda}$ is obtained

where B_y = normal surface induction in gauss

ϕ = remanent flux for tape cross section in maxwells

w = active tape width in cm

λ = wavelength being considered in cm

Results

The outputs of the playback head for the various constant current recorded wavelengths are shown in Figure 2. Corresponding outputs for the nonmagnetic loop are shown in Figure 3. The 3 db down point occurred at a wavelength of 24.2 mils from which the depth dimension of the loop was determined to be 4.6 mils.

The normal surface induction for a wavelength of 4.7 mils was then calculated to be 59.4 gauss peak.

From the search coil measurement the remanent magnetization in a single cross section of tape was found to be 0.256 maxwells. The corresponding surface induction for a signal wavelength of 4.7 mils was calculated to be 60 gauss peak. The surface induction values determined by the two methods were therefore in close agreement. The playback head could then be considered to be calibrated for the 4.7 mils wavelength as having a sensitivity of 7.5 microvolts rms output per gauss of peak surface induction or 94.0 gauss per millivolt output.

The results of measurements made for the various signal and bias levels are shown in Table 1. The agreement between surface induction measurements obtained from the calibrated head and the search coil indicate that, at least for an oriented tape, there is little difference in results for the two methods of measurement.

It should be noted that the search coil only measures the longitudinal component of magnetization whereas the playback head output

Table I

Measured Normal Surface Induction for Various Recording and Bias Levels

Bias Level Percent of Optimum Level	Record Level Percent of Maximum Level	Playback Head Output, rms (millivolts)	Peak Surface Induction Gauss	
			Head Cal.	Search Coil
100	100	.447	59.4	60
100	71.4	.316	42.7	45
100	50	.238	32.2	32.2
100	37	.167	22.5	22.4
137	100	.5	67.5	68.7
70	100	.399	53.5	54.5

is a function of both longitudinal and perpendicular components. For the oriented tape it might be expected that the major magnetization component will be in the oriented longitudinal direction. If an appreciable portion of magnetization were perpendicular, as would be expected from higher bias levels, the results obtained for the search coil would be lower than those for the calibrated head.

Surface Induction at Short Wavelengths

The short gap method is often used to obtain a measure of relative surface induction as a function of wavelength. This relative calibration together with the absolute measurement could be used to obtain an absolute measure of surface induction for all wavelengths. For the extremely short wavelengths being investigated and the 1 micron gap reproduce head employed in these experiments, it was impossible to detect a null point resulting from the gap effect. It was therefore attempted to use the magnetic loop to measure surface induction at short wavelengths.

The normal surface induction for a recorded signal on a tape is defined as the flux density appearing at the surface of the tape in a direction normal to the surface when the tape is suspended in a material (air) of unit permeability. For a recording of constant magnetization in the longitudinal direction at various wavelengths and wavefronts parallel to the tape cross section, losses have been adequately analyzed. When using a magnetic recording head, four wavelength dependent losses are encountered in the playback process.

1. Azimuth
2. Gap length effect
3. Spacing
4. Tape thickness

Azimuth losses are usually small enough to be neglected for carefully fabricated heads. The 1 micron gap length of the playback head and the nonmagnetic loop used in these experiments would introduce only a minor loss at the wavelength of interest. Although spacing loss was held to a minimum by careful polishing of tape and head surfaces, an appreciable spacing loss could be expected.

Tape thickness loss, sometimes referred to as self demagnetization loss was expected to be the predominant factor in the short wavelength response.

On examining the curve of Figure 3, it should be noted that at short wavelengths the curve changes shape at two points, one at a wavelength of about 4 mils and another at a wavelength of about 1.5 mils. It should be realized that for perfect recording and zero playback losses the output of the nonmagnetic loop should be constant for wavelengths appreciably shorter than the "d" dimension of the loop. The inflection points noted should bear a clue to the nature of losses which account for the fall off in output at short wavelengths.

If one were to examine curves of spacing loss and thickness loss as available in the literature, it would be observed that as wavelength decreases, a small loss is gradually encountered. Beyond a particular wavelength the loss increases at a greater rate. It is considered that this change of rate of loss is connected with the change of shape noted in the curve of Figure 3. The first shape change was considered as resulting from one loss, the second by another loss. Thus, if the curve were corrected by the first loss, a flattening of the curve should be achieved in the region between the two points of inflection. Correction by the second loss would then flatten the curve beyond the second point of inflection.

By trial and error, it was found that only thickness loss and one particular value of tape thickness would provide a thickness loss correction which would flatten the curve between the two inflection points without causing the output to exceed that in the original flat region (overcorrection) or cause undesirable irregularities in the curve. This value of thickness was found to be 0.25 mils which agrees with that judged previously to be the effective tape coating thickness. In Figure 4 the output curve for the nonmagnetic loop is shown corrected for this thickness loss.

Next, by trial and error, a value of spacing was found to explain the departure of the thickness loss corrected curve from a flat output. A spacing loss of 0.075 mils was found to best account for this departure. Corrections for this spacing and also a "b" dimension of 0.040 mils, the thickness of the platinum loop, are also shown in Figure 4. The departure of the final corrected curve from flat output was considered to be due to recording losses. Thus, the recording, at least for those layers of tape coating thickness contributing to playback output, could be considered to be perfect to a wavelength of about 0.75 mils. It is interesting

to note that this is the wavelength at which bias was adjusted to give maximum output. At the shorter wavelengths, the recording losses are in close agreement with the difference in short wavelength output obtained under the optimum bias condition of the experiment and that obtained for a bias which is adjusted for peak output for each short wavelength signal.

Under the definition of normal surface induction given, two corrections should be made to the loop output. One is the loss associated with the "b" dimension. The other is the spacing between the loop and tape surfaces. The recording and thickness losses are considered as tape parameters. Actually, the spacing loss can also be considered as a tape parameter. If sufficient care is exercized in the final polish of the loop, almost all of the spacing between loop and tape will be due to the surface roughness of the tape. Should a well polished conventional magnetic head be substituted for the loop, the same spacing loss might be expected.

Then the surface induction for the recorded tape could be calculated directly from the loop output as corrected for the "b" dimension, the equivalent of gap loss. Unfortunately, because the loop output is limited by thickness and spacing losses, at wavelengths less than 0.375 mils, the output was too low to obtain significant measurement. However, the surface induction was easily computed at longer wavelengths.

Analysis of Reproducing Head Results

With the knowledge of the effective magnetization recorded inside the tape coating and tape parameters such as coating thickness and surface roughness, it should be possible to obtain a measure of the output to be expected when the tape is reproduced with a conventional head and compare this with the actual output measured.

For a perfect recording and a perfect reproducing system the output of the reproduce head should rise at 6 db per octave. The recorded level at short wavelengths is less than perfect as previously indicated. The 6 db per octave rising curve should then be corrected for the difference in output for optimum biased and peaked bias conditions. The next loss to be considered is the gap loss. The effective gap length for the reproduce head used was judged to be 0.044 mils. Since the loss due to gap length is minor compared to other losses, appreciable error can be tolerated in this judgment. If the losses used previously for spacing and tape thickness are added, the output determined for the head turns out to be much less than that actually measured as shown in Figure 5. This is explained by the high permeability of the reproducing head core acting as a magnetic keeper on the magnetized elements in the recorded tape. This keeping action tends to reduce the self demagnetization or thickness loss. This keeping effect acts through the entire coating depth for long wavelengths and on only the surface layers of the tape at short wavelengths. Thus at long wavelengths where the full depth is affected and extremely short wavelengths where only surface layers contribute to output there should be zero effective thickness loss. For medium wavelengths where an appreciable percentage of the coating thickness contributes to output, a demagnetizing factor still exists for the deeper contributing layers, some thickness loss should still be evident. This behavior is noted in Figure 5 where it is seen that the thickness loss for the measured curve is essentially eliminated for long and very short wavelengths but appreciable at medium wavelengths.

A combined expression including the keeping action for spacing and thickness losses is available in the literature[4] for cases of both longitudinal and perpendicular magnetization of the tape. For either case, the expressions could not be used to explain the differences in the derived and measured outputs. Possibly the recorded magnetization contained both longitudinal and perpendicular components. The results shown indicate that the short gap method can be used for measuring relative surface induction only when thickness and spacing losses are negligible. The fallacy of a standard reproduce head is also indicated.

If a space were intentionally added between the tape and the reproducing head, the keeping action could be made negligible. This was done by placing a 0.25 mil thick sheet of clear polyester over the reproduce head and remeasuring the output from the recorded tape. The measured output is compared with that determined from the new losses in Figure 6. Agreement between the predicted and measured output is good thus confirming that the keeping action was responsible for the previous difference.

In Figure 7, the measured relative surface induction for the tape is plotted as a function of wavelength. The dotted line represents an extension of the curve by subtracting recording losses, spacing losses and thickness losses as determined herein.

The reproduce head might be calibrated in terms of millivolts output per gauss of surface induction but except for long wavelengths where thickness losses are negligible such a calibration would be misleading. The calibration would be valid for a single variety of tape but since the keeping action reflected in the head output is reported to be a function of tape permeability equal surface induction in different tape coatings can give radically different readings. On the other hand, the definition of surface induction may require alteration. Instead of referring to a unit permeability medium it may be better to refer to the flux generated in a medium of permeability at least 1000 times that of the tape. This would be more realistic since one would be more likely to require the flux available to a high permeability reproduce head than to a medium of unit permeability.

Conclusions

Surface induction measurements obtained by the nonmagnetic loop and the search coil measurements provide essentially equal results for a long wavelength signal.

A carefully fabricated nonmagnetic loop can be used to measure the normal surface induction of magnetic recordings to a wavelength as low as 0.375 mils. If the cross sectional dimensions of the loop are correctly proportioned to the tape coating geometry and recorded wavelengths curves of loop output vs wavelength for a constant current recording can be used to separate losses. These losses can then be used to extend the surface induction measurement to shorter wavelengths.

Acknowledgement

The assistance of Mr. Charles Perone in obtaining measurements and Dr. Walter Guckenburg in performing calculations is gratefully acknowledged.

References

1. Robert Schwartz, Sheldon Wilpon and Frank A. Comerci, "Absolute Measurement of Signal Strength on Magnetic Recordings," J. SMPTE, 64; 1-5 Jan. 1955.

2. O. Schmidbauer, "Zur Bestimmung der Magnetisierung auf Tonband," Electronic Review, No. 10, 1957.

3. Irving Levine and Eric D. Daniel, "Magnetic Properties of Magnetic Recording Tape," J. Acoust. Soc. Am., 32, 1-15 Jan. 1960.

4. W. K. Westmijze, "Studies on Magnetic Recording" Phillips Res. Rep. 8 245-269, 1953.

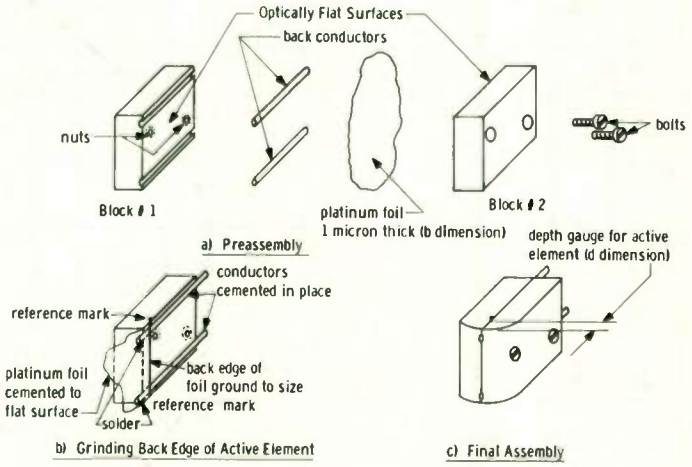


Fig. 1. Construction of nonmagnetic loop.

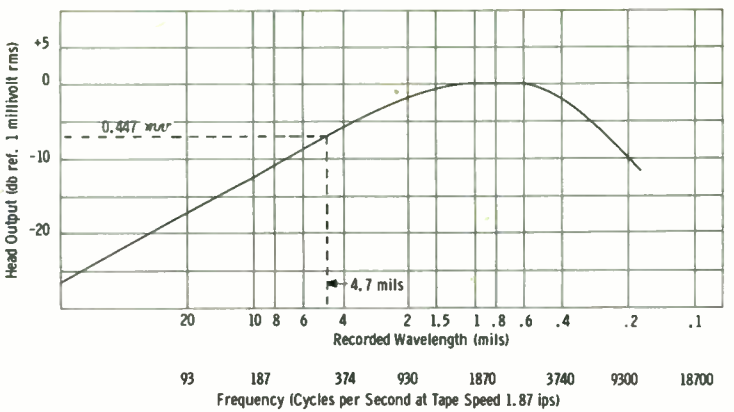


Fig. 2. Playback head output vs. recorded wavelength.

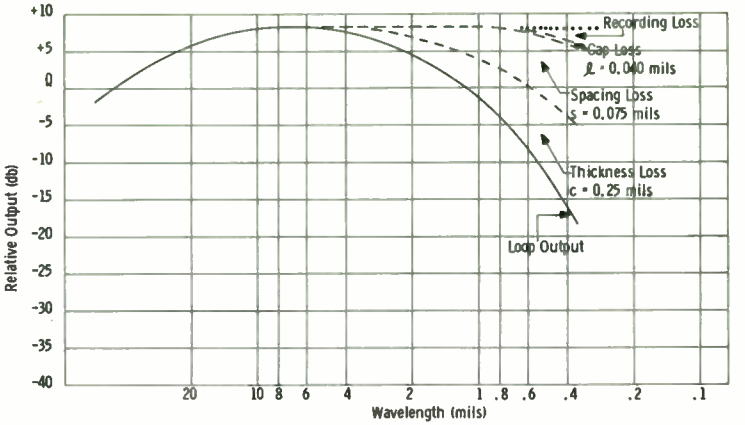


Fig. 4. Loss separation for nonmagnetic loop.

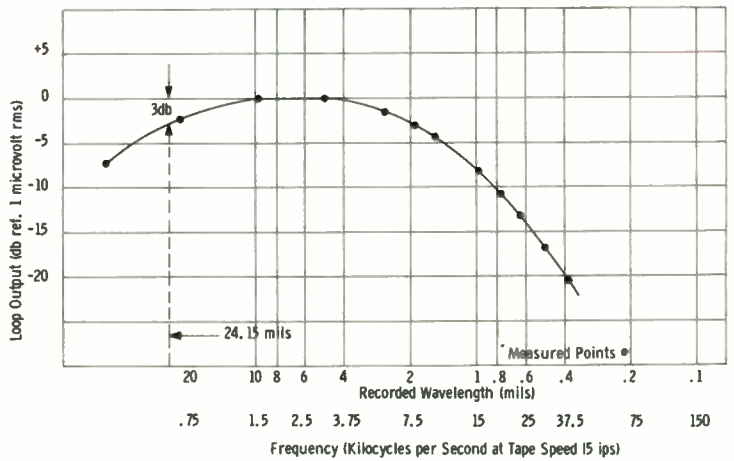


Fig. 3. Nonmagnetic loop output vs. recorded wavelength.

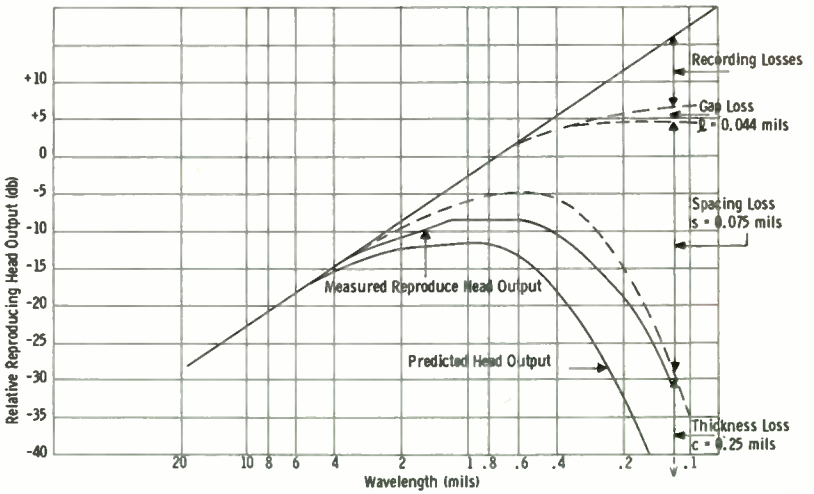


Fig. 5. Loss separation for reproducing head.

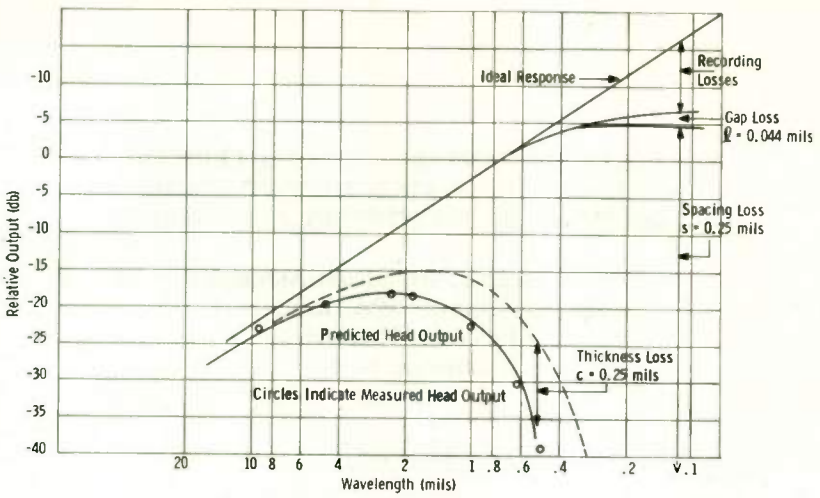


Fig. 6. Loss separation for intentional head to tape spacing intentional spacing 0.25 mil.

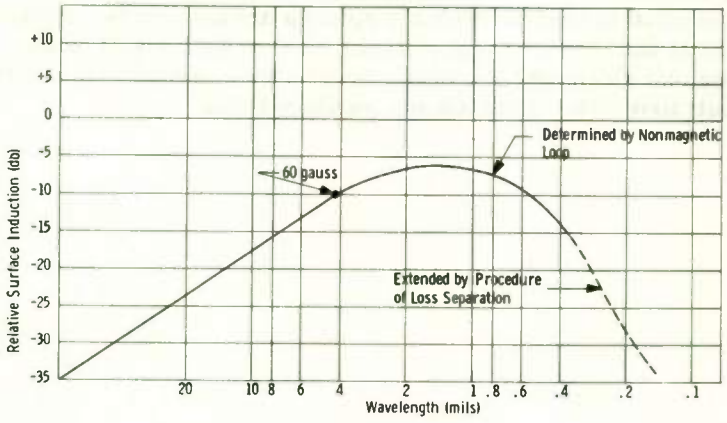


Fig. 7. Relative surface induction vs. wavelength for tape recorded at constant current tape 150 mil wide 0.35 mil thick.

TELEVISION ANTENNA CHARACTERISTICS WHEN IN CLOSE PROXIMITY TO OTHER ANTENNAS OR SUPPORTING STRUCTURES

M. S. Siukola, Broadcast Antennas,
Engrg. Section, RCA, Industrial
Electronic Products,
Camden, N. J.

Abstract

Larger communities are served with several broadcast stations. For both technical and economic reasons it is often desirable to locate the transmitting plants in the same geographical area. With television broadcasting especially, this leads to utilization of multiple antenna arrangements such as antenna farms or even common structural facilities in the form of candelabras, stacked arrays, and interlaced or common antennas.

In this paper the effects of other antennas or supporting structures upon the characteristics of VHF and UHF television antennas are discussed.

Both variations in horizontal and vertical patterns and in impedances as well as decoupling between the antennas are considered in various multiple antenna arrangements. Mathematical and experimental methods used in analysis of candelabra and sidemounted antenna installations are described and illustrated with calculated and measured data.

DIRECTIONAL PATTERNS FROM ELECTRICALLY AND MECHANICALLY TILTED ANTENNAS

R. E. Fisk
General Electric Company
Technical Products Operation
Syracuse, New York

Summary

A directional horizontal plane radiation pattern results from an omnidirectional TV transmitting antenna which has been modified for a combination of electrical beam tilt and mechanical tilt to the antenna mast.

A variety of patterns obtainable by tilted beam techniques is presented along with design considerations necessary for stable performance.

Vertical beam characteristics, power gain values, and pattern control of suitable antenna types are discussed with their application to special coverage requirements.

Introduction

Most Broadcasters and their Engineering Consultants will agree that for every particular TV tower location, there is theoretically an "ideal" antenna that would make the most effective use of the available rf power.

The antenna, as a distributor of rf energy, is the one element in the link of TV Broadcast Transmission that can be adjusted to fit the market area.

In most instances, an omnidirectional horizontal pattern with vertical beam contouring will effectively cover the market. Quite often it is either desirable or mandatory that the radiation characteristic be modified.

It is the purpose of this paper to discuss the control of an antenna's horizontal plane radiation pattern by means of electrical and mechanical beam tilt.

The technique described of electrically and mechanically tilting the antenna beam is a simple method of achieving a desired pattern.

Past Use of Technique

A combination of electrical and mechanical vertical beam tilt has been occasionally used by TV Broadcasters since the advent of relatively high gain and hence narrow beam antennas. Usually it was used to increase the close-in signal in a certain general direction or to move a null in the vertical beam out of a populated area.

Present VHF drop-in applications have increased interest in relatively simple methods of achieving a wide range of horizontal field patterns. For many of these, a tilted beam would seem a means of getting a desired signal reduction in a specific direction, using an antenna of a standard type.

In the case of common ownership of two TV stations, it is sometimes necessary to limit radiation in the direction of the second station so that the two stations do not serve essentially the same area. (FCC Rules, Section 3.636.) In such a case, an improvement in the facilities of one station, such as a higher tower, may mean that a directional pattern must be used to limit the signal toward the second station. Thus, WSPD-TV, Channel 13, in Toledo, Ohio, a Storer Broadcast Company Station, limits overlapping the coverage of WJBK-TV, Channel 2, their station in Detroit by using electrical and mechanical beam tilt.

Vertical Beam Tilt Control

An antenna which has a feed system that permits phasing changes to the individual bays or sections can be readily electrically tilted. The General Electric VHF Helical Antenna shown in Figure 1, for example, is easily adjusted for a wide range of tilt values. Figure 2 shows the feed system used for this helical antenna. For a limited range of tilt values, the electrical tilt may be adjusted from the transmitter room by means of the phasing loop.

Vertical radiation patterns calculated from measured single bay data on the helical antenna exhibiting various values of electrical beam tilt are shown in Figures 3 to 5. The vertical patterns were calculated using a 7090 Digital Computer, and an automatic plotter facility. Notice how use of progressive phasing keeps the contoured vertical pattern relatively constant with the change in electrical beam tilt. In addition, the power gain does not change greatly over the range of from 0 to 2.8° beam tilt.

Horizontal Plane Radiation

The normal horizontal field pattern of a 3 bay VHF Helical Antenna is nearly circular. A typical measured pattern deviates less than ± 1 db from circularity.

Because of this, the horizontal plane pattern with mechanical tilt may be calculated from the resultant vertical pattern angle in the horizontal plane for any particular azimuth angle. See Figure 6.

FCC rules require that the radiated power above the horizontal plane shall not exceed the effective radiated power in a particular horizontal direction in the same vertical plane. (FCC Rules, 3.614(b)(4)). For this reason the mechanical tilt value must be equal or less than the value of electrical beam tilt used. Horizontal plane radiation patterns for several combinations of electrical and mechanical tilt are shown in Figure 7 to 9.

The horizontal plane pattern shape resulting from a tilted circular pattern is shaped roughly like a cardiod. If the antenna has appreciable non-circularity, the horizontal pattern will be somewhat modified. The vertical pattern should be corrected in accordance with the horizontal pattern amplitude for any particular azimuth value.

A batwing antenna having about ±2 db of circularity deviation when tilted in the direction of one of its nulls will have a larger degree of horizontal directivity than a near circular antenna with a smaller tilt value.

Relationship Between Antenna Height and Protection Value

The tilted beam arrangement is best suited to radiation center heights which are 1000 feet or more above the average terrain. Assuming 4/3 earth's radius in such cases the actual signal toward the protected city will be at a depression angle below the horizontal equal to approximately

$$\theta = \sqrt{0.0154\ H_t}\ ,$$

where H_t is the height in feet of the radiation center above average terrain in the direction of the protected area.

Using 1500 ft. as a reasonable value of H_t, the signal we are concerned with will be at a depression angle of approximately 0.6°below the horizontal. To illustrate this, one horizontal pattern is shown along with the field pattern at this depression angle.

Effect of Antenna Sway

The upper side slope of the vertical beam at the depression angle toward the protected market will determine the change in the protection value with antenna deflection. From this standpoint, moderate gain values for example, 15 or less, reduce sensitivity to tower sway.

Conversely, relatively high gain antennas that are fairly flexible should be avoided where it is desired to control a definite signal level in the supressed direction.

To illustrate this point deflection calculations were made for the case of a three bay channel 12 helical antenna with 1.9° electrical and 1.7° mechanical tilt. This antenna has a power gain ratio of 11.5 and a signal supression in the tilt direction of 11 db. The elastic curve of the antenna mast for a ten pound per square foot wind load results in a beam tilt change of 0.19°. The corresponding change in signal level in the tilt direction is +1.4 db to -1.9 db. Some additional shift would be contributed by the supporting structure but conservative design will keep it to a comparatively small value.

New Applications

A tilted antenna installation is now in process at the Taft Broadcasting Company Station WKRC-TV, Channel 12, Cincinnati, Ohio. This installation will use a three bay helical antenna tilted 1.7° toward the Taft Station WKYT, channel 27, in Lexington, Kentucky. Patterns are shown in Figures 10 and 11. In this case, a supression value of about 8 db is desired to limit extending the present Grade B contour closer to Lexington, Kentucky. At the same time the antenna will provide a very high level close-in signal in the supression direction. Since the station is located in metropolitan Cincinnati, the strong close-in signal is desirable. A conventional directional antenna would reduce the close-in signal toward Lexington compared to other directions, while the tilted antenna increases it.

The patterns of the WKRC-TV antenna will be carefully measured before installation. It is expected that many field measurements will be made of the signal toward Lexington so that the observed supression can be compared with the theoretical value.

Conclusions

1. A tilted antenna makes possible a wide range of directional horizontal patterns without extensive changes in the basic antenna.

2. Antennas which are readily adjusted for beam tilt and which have excellent mechanical stability are well suited for this technique.

3. For moderate values of signal supression the tilted beam concept would seem useful where a second station is to be protected as required for VHF drop-in service.

4. Measurements of an installed antenna such as WKRC-TV will provide additional field test information on such applications.

Fig. 1.

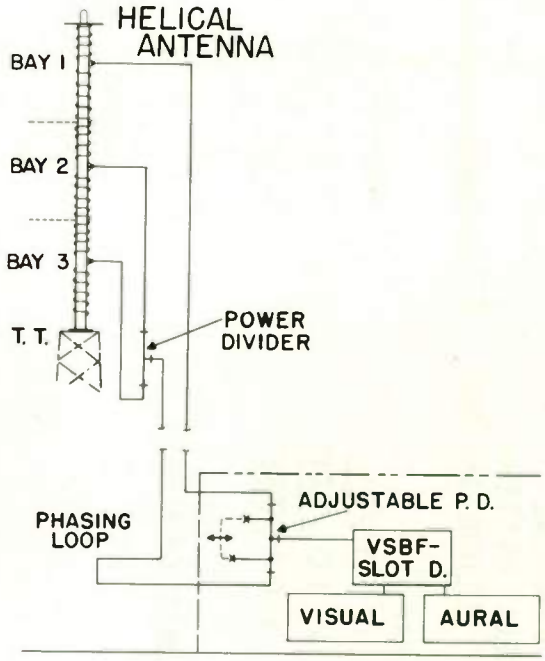


Fig. 2.

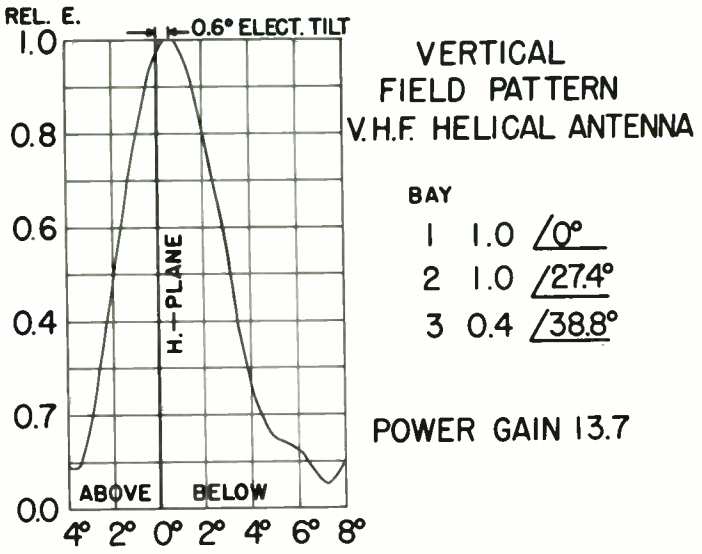


Fig. 3.

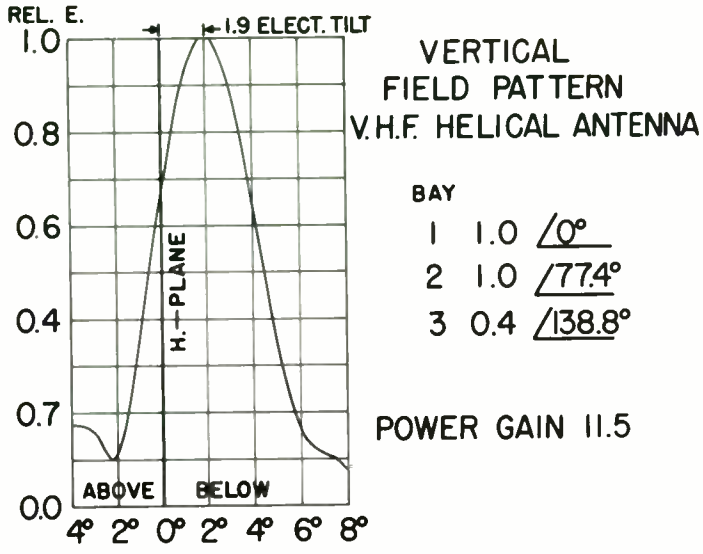


Fig. 4.

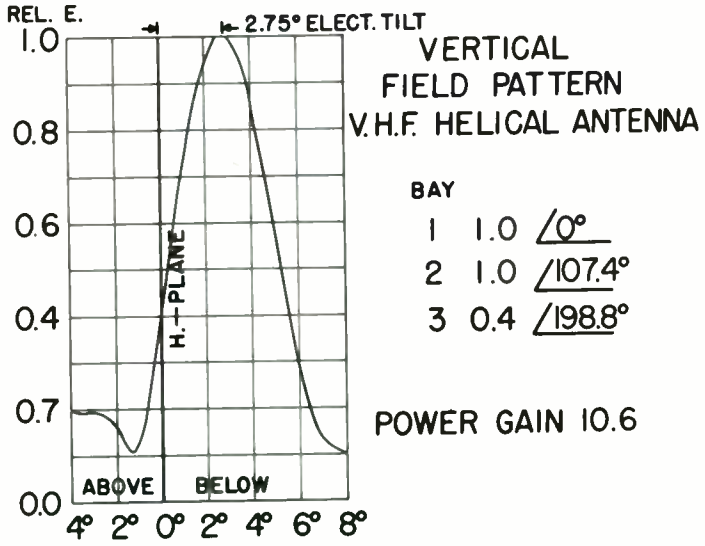


Fig. 5.

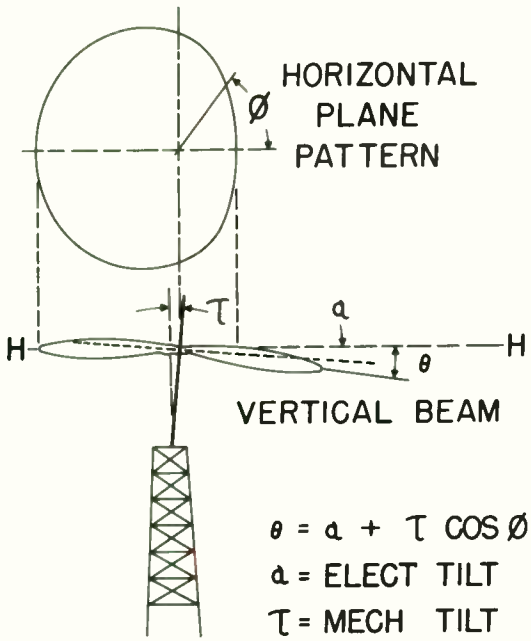


Fig. 6.

DIRECTIONAL HORIZONTAL FIELD PATTERN

ELECT. BEAM TILT 1.6° DOWN
MECHANICAL TILT 1.6° NORTH

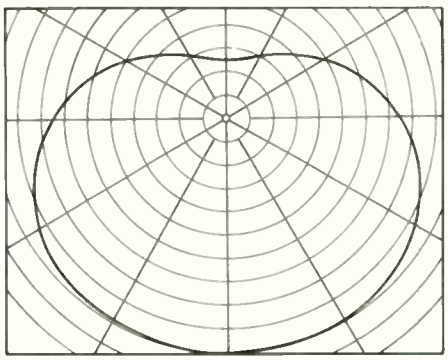


Fig. 8.

DIRECTIONAL HORIZONTAL FIELD PATTERN

ELECT. BEAM TILT 1.4° DOWN
MECHANICAL TILT 1.4° NORTH

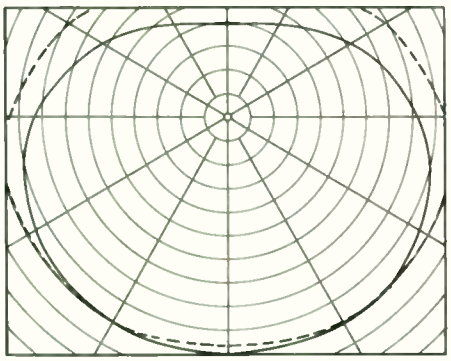


Fig. 7.

DIRECTIONAL HORIZONTAL FIELD PATTERN

ELECT. BEAM TILT 1.9° DOWN
MECHANICAL TILT 1.9° NORTH

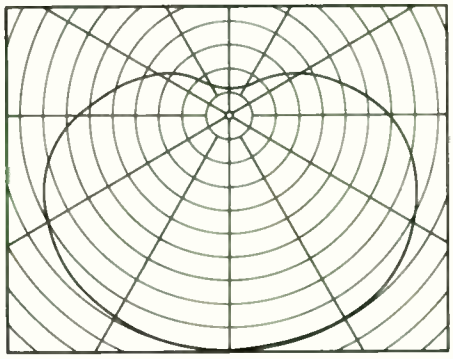


Fig. 9.

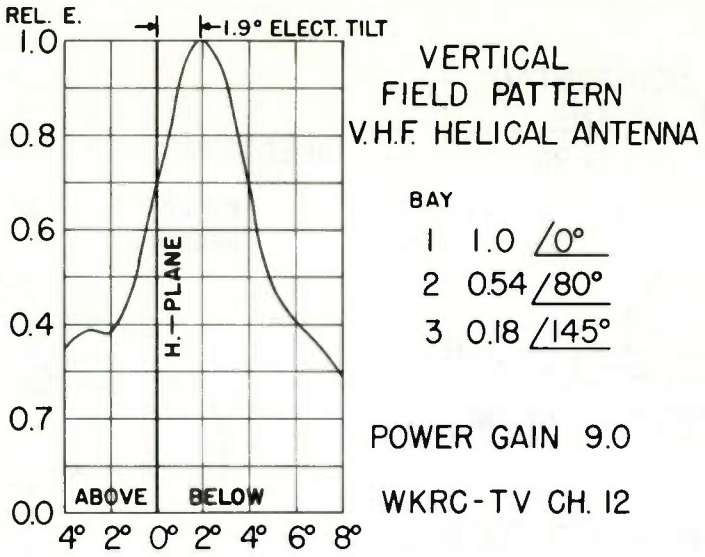

REL. E.
1.0
0.8
0.6
0.4
0.7
0.0
1.9° ELECT. TILT
H.—PLANE
ABOVE
BELOW
4° 2° 0° 2° 4° 6° 8°
VERTICAL
FIELD PATTERN
V.H.F. HELICAL ANTENNA
BAY
1 1.0 /0°
2 0.54 /80°
3 0.18 /145°
POWER GAIN 9.0
WKRC-TV CH. 12

Fig. 10.

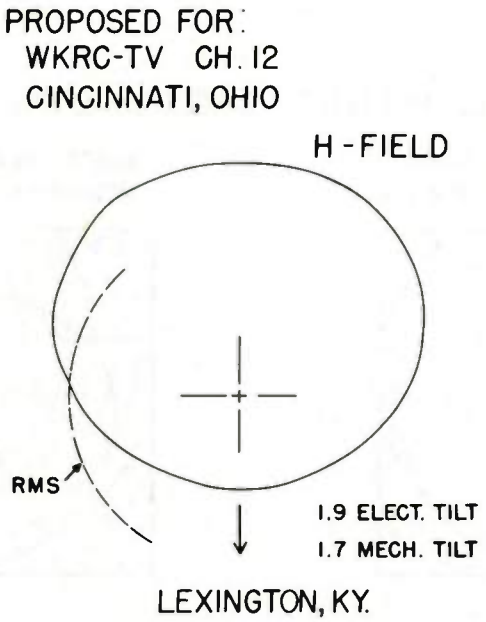

PROPOSED FOR:
WKRC-TV CH. 12
CINCINNATI, OHIO
H-FIELD
RMS
1.9 ELECT. TILT
1.7 MECH. TILT
LEXINGTON, KY.

Fig. 11.

EARLY RETURNS FROM WUHF

Arnold G. Skrivseth
UHF-TV Project Chief
Office of the Chief Engineer
Federal Communications Commission
Washington 25, D. C.

Summary

This paper is a sequel to one given on October 6, 1961, at the Washington, D. C. meeting of the Professional Group on Broadcasting and published in the December, 1961 transactions of that group. The present status of the Commission's New York UHF-TV Project is described along with comments and tentative conclusions based on early results.

- - - - - - - - - - - - - -

The experimental television broadcast station operated by the Federal Communications Commission under call letters WUHF has been in operation from the Empire State Building in New York City since October, 1961. At that time daily transmissions were started using a directional antenna. Since November 26, 1961, an omnidirectional antenna has been used.

During the directional antenna operation in October and November, polarization was alternated between circular and horizontal polarization, on a half-hourly schedule. Transmissions during the first half of each hour were on horizontal polarization. During the second half of each hour, circularly polarized signals were radiated. The changeover between polarizations required about 2 minutes. The purpose of this alternating form of operation was to permit direct comparison of the two modes of polarization as radiated by the same antenna.

The antenna pattern of the directional transmitting antenna was measured by Melpar, Inc. prior to shipping this antenna to New York. Since the back radiation of this antenna is very low, it seems reasonable to assume that its radiation pattern when mounted in a window of the 80th floor of the Empire State Building remains the same as when tested on an antenna range. As a result, in addition to use of this antenna for comparison of circular with horizontal polarization, measurements have been made to compare the radiation of the main antenna with that of the horn antenna to serve as a check on the pattern of the main antenna.

Measurements and observations were made on the signals radiated from the directional antenna by 11 crews of two men each who are employees of Jerrold Electronics Corporation under contract with the Commission. These crews were equipped with a portable television receiver, a portable field strength meter, a corner reflector antenna and right hand and left hand helical antennas. The gain of the corner reflector antennas is nearly the same as that of the helical antennas when each is receiving the polarization for which it is designed.

The transmitting antenna gain on horizontal polarization is nearly identical to its gain on circular polarization, since the radiation pattern is nearly identical, provided, of course, that one measures gain with the appropriately polarized receiving antenna, i.e. a horizontal antenna for horizontal polarization and a right hand circularly polarized antenna for circular polarization. With the equipment thus selected, one can compare directly the results of reception on one polarization with that of the other without introducing a number of questionable compensating factors. For those interested in numbers, the gain of this antenna with reference to a dipole is about 14.5 db for horizontal polarization at 573 Mc. For circular polarization its gain is about 3.5 db less. However, if a circularly polarized receiving antenna is used then the gains are within 0.5 db of each other.

With the beamwidth of this antenna, which is about 30 degrees to the half power points, there is a limited area centered on a bearing N29°E (which is the direction of uptown Manhattan from the Empire State Building) within which observations and measurements can be made. Within this sector and within 25 miles of the station 170 locations were selected by the U. S. Census Bureau on a statistical basis. Of these 70 locations were accessible to the project. At 35 of these 70 locations, measurements and observations were made with roof-top antennas as well as with indoor antennas. Since this was a very limited sample an additional 128 locations were surveyed which were either at convenient locations within the designated area or were at fire houses, police stations and the like. Also, spot measurements were taken at about 400 locations. Data at both types of locations have been analyzed to determine what differences there might be between the two type of polarization. This analysis has

been done manually but at a later date a more detailed analysis is planned after the data have been reduced to a form suitable for automatic processing. Inasmuch as we are busily occupied with obtaining data on transmissions from the main antenna and in transferring all the data over to punch cards, a complete analysis will be somewhat delayed. However, the analysis we have made to date of these data leads us to conclude that under the circumstances of this experiment, neither circular nor horizontal polarization offers a substantial advantage over the other.

This is not to say that at a particular receiving location that one or the other mode of transmission might not show a recognizable superiority. What is meant is that when all locations which we investigated were considered, we found essentially as many locations where horizontal polarization provided better reception than circular as locations where circular polarization was judged superior. At several of the points in weak signal areas circular polarization seems to be consistanly somewhat better than horizontal polarization. However, this evidence is too limited to be conclusive.

Further clarification of the foregoing comments is necessary to the extent of making a comment on what is meant when saying "better than." For the purpose of this discussion I mean that based on the judgment of the observer, which in this case is a trained observer, a particular picture is better than another when all causes of degradation are considered. In most all cases the degradation is either due to ghosting (which includes smearing) or to thermal noise.

In addition to the data gathered by the Jerrold crews, we are operating a field car to determine in a more conventional manner the extent of service. During the directional antenna phase mobile runs were made at 10 ft. antenna height on a radial in the center of the beam (bearing N29°E) on both horizontal and circular polarization. Similar runs were made on arcs at about 23 miles and 40 miles from the station. Only a portion of these data has been analyzed. The field car is fitted with a pneumatic mast which extends to a 30 foot height. Most of the data taken with this car has been at 30 feet. The technique used is that which was used by TASO and others. A radial is carefully drawn on topographic maps and points marked off at 2 mile intervals. The nearest accessible point to the exact 2 mile point is used for taking data. After a point is selected where a run can be made at a 30 foot height the appropriate antenna is attached the antenna is elevated to 30 feet and the car, with field strength equipment operating, is slowly driven over a distance of at least 100 feet (where possible). Recordings of the field strength are later analysed and the values obtained are tabulated and plotted on a graph.

In the case of the directional transmitting antenna, the car was used for 30 foot measurements over the same radial and arcs as for the 10 foot measurements. However, since roads do not run conveniently in the right places, only infrequently are the points on the radial and the actual road radial run within a reasonably close distance to each other. Data for circular polarization was obtained with a right hand helix, a left hand helix and a corner reflector. On horizontal polarization data was taken only on the corner reflector antenna. In most instances the circular polarization data includes a rotation of the corner reflector about a horizontal axis to obtain information on the polarization of the signal received.

The data from the field car when a directional transmitting antenna was in use have been only partially analyzed. However, the analysis is sufficient to show that here too there is little, if any, difference between circular and horizontal polarization on an overall basis when measured under the conditions specified. Occasionally one does find a point where circular polarization seems to have an appreciable improvement over horizontal polarization. This seems to occur in spots where the direct signal is highly attenuated and a reflected signal is obtained from a direction widely different from that of the station. I can recall only two instances of this out of approximately 75 locations. However, even in this case the reflected signal was only slightly better than that obtained on horizontal polarization. On the other side of the coin it should be stated that there were many locations where horizontal polarization was better than circular polarization by a small amount. However, in most of these locations there was adequate signal in either case. A more careful study of the data will be required before any further conclusions can be drawn. In this connection, I would be interested in obtaining any ideas on what use can be made of the information obtained on the polarization of the incoming signal when circular polarization is being transmitted. This information is definitely interesting but until analyzed by some logical method, not particularly useful.

Perhaps the most that can be said of the comparison of circular and horizontal polarization as measured under the conditions herein is that when a practical broadcasting situation is considered, there is not much to choose between the two, even assuming that it is equally convenient to transmit and receive either polarization. However, there does seem to be evidence to indicate that with circular polarization there might be an advantage in highly attenuated areas if one were to use highly specialized antennas suited to the particular location. Whether or not a carefully engineered receiving set-up would remain useful from day to day and season to season for circular polarization remains a question. Also whether or not the same money spent in an installation for horizontal

polarization would result in an equal picture is also open to question. We have investigated this matter from what we consider to be a practical broadcasting situation. We have not investigated it from a highly scientific point of view. One could argue that we should but it seems more realistic, since the differences appear to be only of academic interest, to leave further studies of this matter to laboratory measurement under controlled conditions. Further analysis of our data may of course result in somewhat different conclusions.

A work of caution seems to be in order, however. Our observations and measurements were made on a directional antenna. Use of an omnidirectional circularly polarized antenna would no doubt produce somewhat different answers. Although one type antenna was proposed for this project which presumably would produce omnidirectional circular polarization the complexity of this arrangement appeared to us to be unreasonable from a reliability point of view. In addition, since wide spaced antennas have considerable scalloping, and the effects of the existing antennas on the building on the polarization are not known, there seemed to be no adequate way to determine what polarization would in fact be radiated from such a complex array in such a complex environment.

Leaving the subject of circular polarization, we now come to the present operation which consists of an omnidirectional antenna mounted with its radiation center about 1330 feet above sea level on the Empire State Building. This is the antenna described by Dr. R. Wayne Masters of Melpar, Inc.

The area of operation of the Jerrold crews now becomes that within a circle of 25 miles radius centered on the Empire State Building. Operation on this antenna commenced on November 26, 1961. Since that date and until the end of February observations and measurements were made at over 500 locations. Installations of receivers were made in over 100 homes. The observations and measurements continue as do the receiver installations at a pace such that by the end of October, 1962 we should approach our design goal of 5000 and 1000 respectively.

Since November 26 our comparison has been of channel 31 versus channels 2 and 7. Preliminary results on the first 100 locations for measurements and observations indicated that there is very little difference, within the 25 mile circle, between the three channels. The slight difference indicated so far appear to favor channel 2 over channel 7 and both channels 2 and 7 over channel 31.

The following table shows the results of preliminary analysis of these 100 locations. It should be kept in mind this is a very limited sample of data and that any conclusions that one might draw from it are subject to considerable question as to their validity when applied to New York City as a whole.

Picture Quality Observations

Picture quality was graded at all locations using original TASO photographs* as a standard for judging degradation due to noise, and using a special scale for judging degradation due to ghosting (described in ANIAC Report). Results of these observations may be tabulated as follows:

(1) Degradation due to Thermal Noise

Percentage of locations where grade is equal or better.

Pict.	Channel 2		Channel 7		Channel 31	
Grade	Indoor	Roof-top	Indoor	Roof-top	Indoor	Roof-
1	49%	94%	43%	91%	43%	89%
2	85	98	85	97	78	97
3	96	100	94	100	87	99
4	97	100	99	100	94	100
5	100	100	100	100	99	100
6	100	100	100	100	100	100

(2) Degradation due to Ghosting

Percentage of locations where grade is equal or better.

Pict.	Channel 2		Channel 7		Channel 31	
Grade	Indoor	Roof-top	Indoor	Roof-top	Indoor	Roof-
1	31%	71%	32%	77%	32%	69%
2	77	96	68	94	68	94
3	95	97	91	97	86	98
4	97	99	96	99	93	99
5	100	100	99	100	98	100
6	100	100	100	100	100	100

*Engineering Aspects of TV Allocations, TASO Report to FCC, March 16, 1959, page 254.

Another group of data has been analyzed on a preliminary basis from the first 100 locations where receivers were installed. The results of this analysis are similar to that shown above in major respects. Here too I must caution you not to come to any hasty conclusions since this is a rather small sample.

Since the Jerrold crews only work out to a radius of 25 miles, we are supplementing this information with field measurements made with our field car. Two radials were completed prior to the end of February. These are in the directions N29°E and N50°E. Work is in progress on radials at 80, 190, 230, 290, 340 and 0 degrees, all expressed as E of true N. No unexpected results have been obtained in these measurements. Or expressed differently, UHF is inferior to VHF at the greater distances in rough terrain. In smooth terrain UHF compares favorably with VHF out to line of sight distances, which in this case is about 50 miles.

Our field car is also equipped with a television receiver of the same type used by the Jerrold measurement crews. Initially pictures were watched at various locations to determine picture quality. However, it was soon determined that if sufficient signal was available there was no difficulty experienced in obtaining an adequate picture. This comment, of course, refers to the regions outside the densely populated areas since the field car has not been operated within the central portion of New York City. In addition, since the antennas used for VHF, which are simple dipoles, are not typical of the type used in the home as is the corner reflector used on UHF, it would not be a fair comparison of picture quality to compare pictures on UHF and VHF. A further factor is that with ordinary ignition suppression as is used on our vehicle, the noise level on VHF is very high. Since it is not typical for a motor vehicle television set to be operated continuously within 30 feet of a television antenna, this also would cause difficulty in rating picture quality. In this connection, it may be of interest to those of you who have not made such field surveys to note that noise levels found along roads present considerable difficulty in recording field strength on VHF whereas on UHF only about half the automobiles and trucks create enough interference to be recorded on the equipment in use. These difficulties, of course, are only experienced at the outer fringes of service.

The field car has performed an additional function in this project and this is to make a check on the performance of the main antenna. Identical mobile runs have been made at an antenna height of 10 feet using the horn antenna in one case and the main antenna in the other. These runs to date have only been made in the area where the horn antenna provides a reasonable signal level. One set of runs has been made on a 23 mile arc (this distance was chosen to roughly coincide with the Tappan Zee bridge) and another set on a 40 mile arc (which is roughly the distance to the Bear Mountain bridge). One cannot expect completely consistent data under the circumstances measured since at 10 feet the field strength received is quite sensitive to the amount of surrounding automobile and truck traffic. This is particularly true of the 23 mile arc since this was run on Interstate Route 287 which is highly travelled. However, when all the known factors have been accounted for such as the horn antenna pattern and the omnidirectional antenna pattern (including its vertical pattern) the pattern of the latter antenna checks with the model measurements within about plus or minus 5 db. This, of course, can only be done over about 40 degrees of arc. However, at least one can say at this time that we have so far no indication that the main antenna is not performing as designed. In fact, since the antennas are displaced about 320 feet vertically from each other the ± 5 db variation could be attributed to this fact alone. Again, at those roof-top locations where clear line-of-sight is obtained to the transmitting antenna the Jerrold crews have been measuring signal levels between 100 and 115 dbu out to about 7 miles. These values check very nicely with the design value of 110 dbu, although we have not yet checked where these locations are in the complex antenna pattern.

INTERLEAVED SOUND
TRANSMISSION WITHIN THE TELEVISION PICTURE

J L Hathaway
National Broadcasting Company
New York, N.Y.

Summary

A system is described for combining sound and picture signals and transmitting them over a single video circuit in such a manner as to provide an emergency sound service for use during failures of the regular television audio facilities. Picture and sound portions of network programs are ordinarily carried by inter-city circuits which differ in apparatus and routings. Although each has been subject to occasional service interruptions, simultaneous failures have been highly unusual. This Interleaved Sound system creates no degradation of picture quality, adds no interference components to the picture signal, and provides recovered sound of adequate quality.

The Problem

Television broadcasting networks are occasionally troubled by failures of intercommecting common carrier circuits, even though advanced, reliable, well maintained equipment is used. Although television sound failures attributable to the common carrier have become less frequent in recent years, a tabulation for 1960 on NBC circuits shows many hundred interruptions lasting longer than 30 seconds. As an example of one, resulting from a switching error, CBS audio was routed on Bob Hope's video over NBC; thereby interrupting several minutes of sound and providing front page newspaper coverage. Sometimes it seems the public is so engrossed with pictures that sound gets little consideration until there is a break in its continuity; then the lack of sound becomes of primary importance.

Picture and sound circuits are normally carried over inter-city facilities which are different in types of apparatus and in routings. One major difference is that video circuits are one-way only, whereas the audio is reversible by means of a series of relays operating in tandem at the many repeaters along the route. Actuation of these relays is sometimes adversely affected by spurious voltages developing along one or more of the many lengths. This and many other factors have caused television sound outages, but it is noteworthy that these have seldom occurred simultaneously with picture circuit failures.

Early Considerations

The need for economical emergency sound facilities has long been recognized by broadcasting networks, in fulfilling their mission of public service. Several years ago consideration was given to development of a system similar to that employed in microwave relaying, wherein a frequency modulated channel is multiplexed above the video frequencies. This seemed expedient and might have provided a high quality circuit if the video spectrum had been slightly curtailed. However, it would not have functioned on circuits which did not transmit the multiplexed channel properly. The decision to avoid using the extremely high end of the video spectrum proved wise, in view of the fact that for reasons of economy broadcasters are now using somewhat narrower inter-city facilities.

Interleaving

A system of Interleaved Sound has now been developed to a state of practical utility, wherein video and emergency audio are transmitted over a single video circuit. In this system, the sound is interleaved within the picture components like a sheet of paper between two of the many pages of a book, while the book remains intact. This means that the band requirement is no greater than for video alone. It also means no added hardships are imposed in handling the video and the audio.

The system of interleaving sound is possible because of the nature of television signals. Many years ago it was found, experimentally and mathematically, that television scanning of a picture produces concentrations (or "clumps" of energy) distributed throughout the video spectrum at harmonics of the horizontal line scanning frequency, each having sideband components at multiples of the field scanning frequency. Between these clumps there is little video energy and it is into one of these "gaps" that sound is inserted. A few of the energy clumps are represented in Figure 1, together with audio signals interleaved between two selected clumps -- in this instance the 113th and the 114th harmonics of line scanning frequency. The gap is rather narrow, especially for pictures including considerable high-contrast finely detailed material; so for this reason, single-sideband, suppressed-carrier transmission is used for the sound. Any other gap in the same general portion of the spectrum could have been chosen and used just as well as that between the 113th and 114th harmonics. Tests showed, however, that video interference between the clumps increases at lower frequencies. In the direction of higher frequencies the interference would have been much lower, if the gaps were not occupied by chrominance components during the transmission of color.

Transmitter & Receiver

A single-sideband sound generator and a sound recovery unit were constructed in 1958 for evaluating the principal of interleaved sound. Figure 2 diagrams these units in block form and is generally self-explanatory. Both the transmitter and receiver are "filter" types, employing unsymmetrical crystal filters to pass only the upper sideband and to give sharp attenuation of frequencies near and below the suppressed carrier frequency. The filters have identical characteristics, with measured frequency response as shown in Figure 3.

Pass range of the crystal filters is centered between the 113th and 114th harmonics of the average frequency of horizontal scanning. This centering results in minimum interference from video into the audio and vice versa. Monochrome television horizontal frequency in the USA averages 15,750 cps, with slight variations caused by changes of the 60 cps power line frequency which is generally used to control the synchronizing generators. Color television employs a horizontal frequency 0.1% lower. Center frequency for the audio energy is determined by multiplying one half the average horizontal frequency by an odd integer. Sound is placed between the harmonics previously mentioned when this integer is 227, thus:

$$f = \tfrac{1}{2} (15{,}750 - 0.05\%) \times 227 = 1{,}786{,}731 \text{ cps.}$$

The carrier should be located approximately 1000 cps lower in frequency, or at 1,785,700 cps.

Temperature regulated quartz crystal oscillators are used in the transmitter and receiver, and a means of readily checking frequency of the receiver against the transmitter is provided. These frequencies must match within about 4 cps, as the recovered sound will otherwise be offset in frequency enough to implant a slightly unnatural characteristic to certain voices and other program content. Operationally, the frequency check and adjustment may be accomplished at any time when the emergency circuit is not in use. A switch at the transmitter applied 120 cps tone modulation which is derived from the 60 cps power frequency. If, at the distant recovery unit, the audio output exceeds the limits 120 $\pm$ 1 cps, the local oscillator should be trimmed. Extreme accuracy is possible in this checking, by connecting an oscilloscope to the receiver output and sweeping the trace with the local 60 cps power frequency. In metropolitan areas, power frequency is generally maintained within 0.1 cps. Under typical operating conditions, where Interleaved Sound and all surrounding apparatus are turned on 24 hours daily, stability of the oscillators is such as to require a frequency check on a weekly basis. The crystal filters and all other components are sufficiently stable to make balancing and gain readjustments unnecessary over months of continuous operation.

Preliminary Results

Laboratory testing of the system was first conducted by adding the single side-band sound to the picture signal and recovering the sound components without any long interconnecting circuit. Results were good enough to warrant trials over actual video circuits, so a test was conducted from Radio City to Brooklyn and return. This test also was very satisfactory, so next a circuit of over 2000 miles from New York to St Louis to Chicago and return was tried; again with equally good results. Finally, as a check on transmission over extreme distances, tests were made on a coast to coast and return common-carrier of more than 7000 miles total. Here again results were good; there was no significant degradation of sound as compared to the original laboratory bench tests conducted without any inter-city circuitry. Furthermore, the video signals suffered only the normal degradation associated with such long-distance transmission.

Control of the level of the Interleaved Sound signals is extremely important, since an excessive level creates interference patterns on the picture while insufficient level causes objectionable buzzing in the recovered sound. For the tests over various long distance circuits, the normal level of sound transmission produced peak to peak voltage measuring one percent of the peak to peak television signal voltage. Figure 4 depicts a video waveform with and without the added sound signals. Under this condition of transmission, no trace of sound interference could be seen on the picture, even on closeup viewing of large high quality monitors. At this transmission level the recovered sound was highly intelligible and adequate for use during short periods of interruption of the regular audio circuit. For prolonged listening, however, an improved signal-to-noise ratio was highly desirable; so the plan was to have the receiving end of the circuit initiate a telephone call to the transmitting end as soon as possible after an interruption occurred and the emergency circuit had been placed in use. The call would bring about a fourfold increase of transmitted sound level and receiver sensitivity would be dropped 4 to 1. This resulted in a recovered sound signal-to-noise ratio improvement of 12 db. Although this caused 4% modulation of the television signal by the sound, interference into the reproduced picture was insignificant and unnoticeable in ordinary television viewing.

A 30 day test was conducted, with FCC authorization, in the Fall of 1960 from New York, with a sound recovery unit located at Washington, DC. Sound signals were added to all programs originating in or passing through the Radio City, New York control center. Normally 1% level prevailed but on many occasions the 4% level was used to simulate emergency conditions. AT&T Long Lines personnel as well as those of all NBC affiliated stations were alerted prior to the test; however, some days after its inauguration,

reports indicated that many had not only failed to see any sign of interference to the pictures, but were wondering why the test had not started as scheduled! During the entire month there was not one comment from the public which could in any way relate to interference from the added sound signal; furthermore, the network affiliates reported no interference. A demonstration was given to FCC members and staff on the final day of testing.

Improved System

Success of the 1960 test lead to the trial of another development. This has proven well worthwhile and results in three major advantage over the earlier system:

1) The signal-to-noise ratio of the received signal is always satisfactory for immediate and prolonged emergency use and the level need not be increased above the normal 1% value.

2) Satisfactory signal-to-noise ratio is obtained at lower sound transmission level, so there is never the slightest vestige of interference to the picture.

3) Elimination of the level change required when a sound interruption occurs at a distant city simplified operation as compared to the two-level system.

The only change apparatus-wise for realizing these advantages is in the addition of a special filter to clear the channel for the subsequent injection of Interleaved Sound signals. A double-notch filter eliminates most of the video components which would ordinarly occur at relatively low levels between two adjacent clumps of video energy. In the early system, practically all of the adverse noise in the recovered sound was created by video crossover; therefore a reduction of these crossover components reduces the noise to a degree dependent on the filter's attenuation throughout the channel. A block diagram of the over-all transmitting and receiving system for video, regular audio, and emergency audio is given in Figure 5.

Notch Filter

A suitable filter must be sufficiently selective to reject the desired band of frequencies without material attenuation or phase shift at the frequencies of the adjacent video clumps. These criteria are met in the present filter by using LC-tuned circuits arranged to cancel video at two frequencies. The over-all filter, with isolation amplifier for over-all unity gain in a 75 ohm circuit, is shown in block form in Figure 6. The LC-tuned circuits have a Q of around 150 at the tuned frequency of approximately 1.78 mc. Each of these is provided with an amplitude and phase control (as well as tuning control) so that after passing through a cathode follower, cancellation of the video signal may be achieved at the resonant frequency. The two LC circuits are slightly staggered in tuning to give the response shown in Figure 7. This curve is plotted with greatly expanded scale of frequencies around the sound channel in Figure 8. Although the unit does not represent the ultimate in performance for this application, it has proven relatively stable and easily controlled. Furthermore, it provides a subjective improvement of at least 12 db in the recovered sound signal-to-noise. A preferred filter design would utilize a number of crystals which would give four or more rejection frequencies. Such a filter should cover a wider band of frequencies having at least 20 db of attenuation but a narrower band having no significant attenuation or phase shift. This would result in improved signal-to-noise ratio of the recovered sound without introduction of video "ringing."

Performance & Status

The over-all system described and presently in use as an emergency sound circuit covers the audio frequency band from 100 cps through 4300 cps, flat within $\pm$ 2 db. Harmonic distortion is below 3% RMS. Signal-to-noise of the recovered sound is a changing factor, depending on the nature of the video. In general, measured RMS noise is between 27 db and 45 db below RMS program; average video produces a figure of approximately 36 db. This is substantially the same on coast-to-coast circuits as on very short circuits.

Interleaved Sound transmitted at the specified level and radiated by a broadcaster produces no detectable change or deterioration of received picture quality and no co-channel or adjacent channel interference. Furthermore, the level of transmission is so low that the sound signal which is added to video is far below the system noise limit specified by the FCC. Thus, the broadcaster's service to the public is never deteriorated by the existence of emergency sound, but whenever there is a failure of the regular audio circuit, this system makes all of the difference between having and not having sound.

An agreement has been negotiated giving the common carrier ownership of the experimental equipment, and the network is now operating this over the New York to Hollywood circuit. The operation has been authorized by the FCC for a period of one year. Presently the American Telephone & Telegraph Company is testing other systems of transmitting sound together with video over a single circuit. They hope to evaluate the various possible systems and to carry to completion the development of whichever proves the most practical.

Acknowledgment

The author appreciates the encouragement and help received in pursuit of this development from Andrew Hammerschmidt, Theodore Kuron, George Nixon, William Trevarthen and James Wilson.

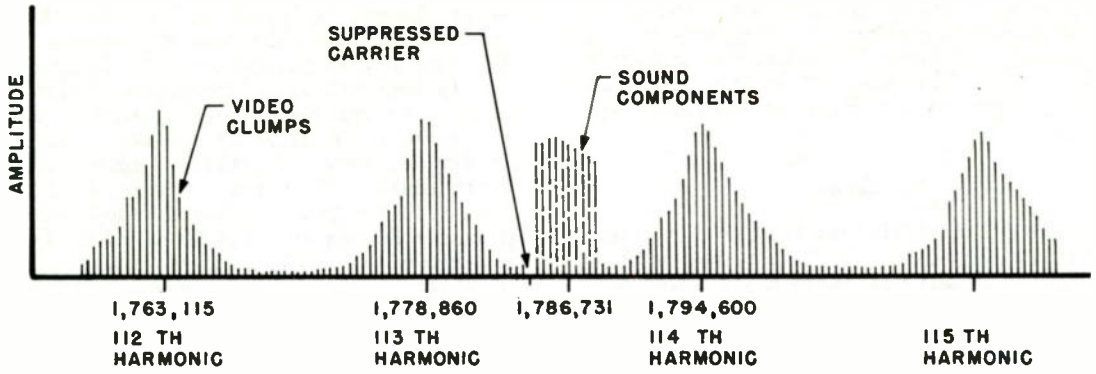
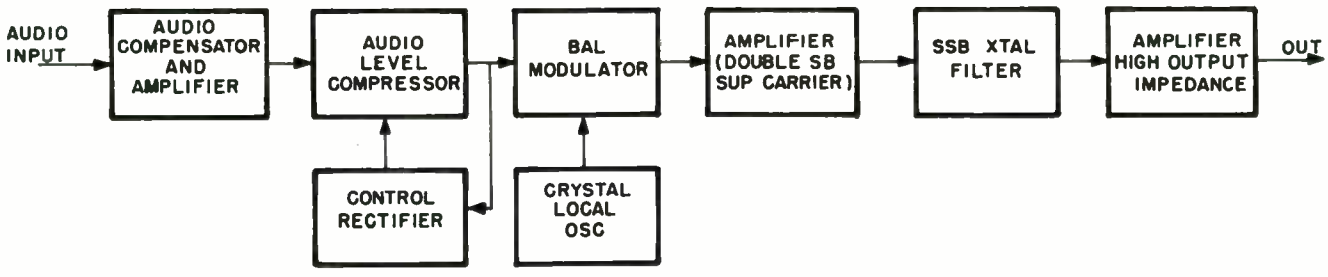
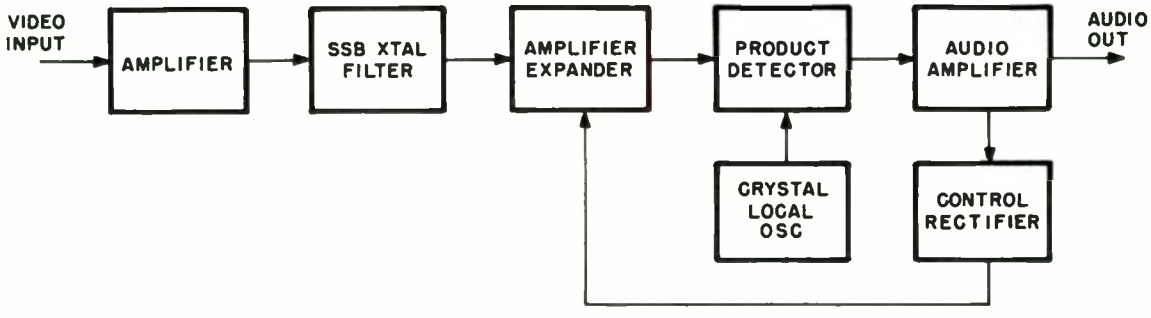


Fig. 1. Video energy clumps and interleaved sound.



EMERGENCY SOUND GENERATOR



EMERGENCY SOUND RECEIVER

Fig. 2. Single-side band block diagram.

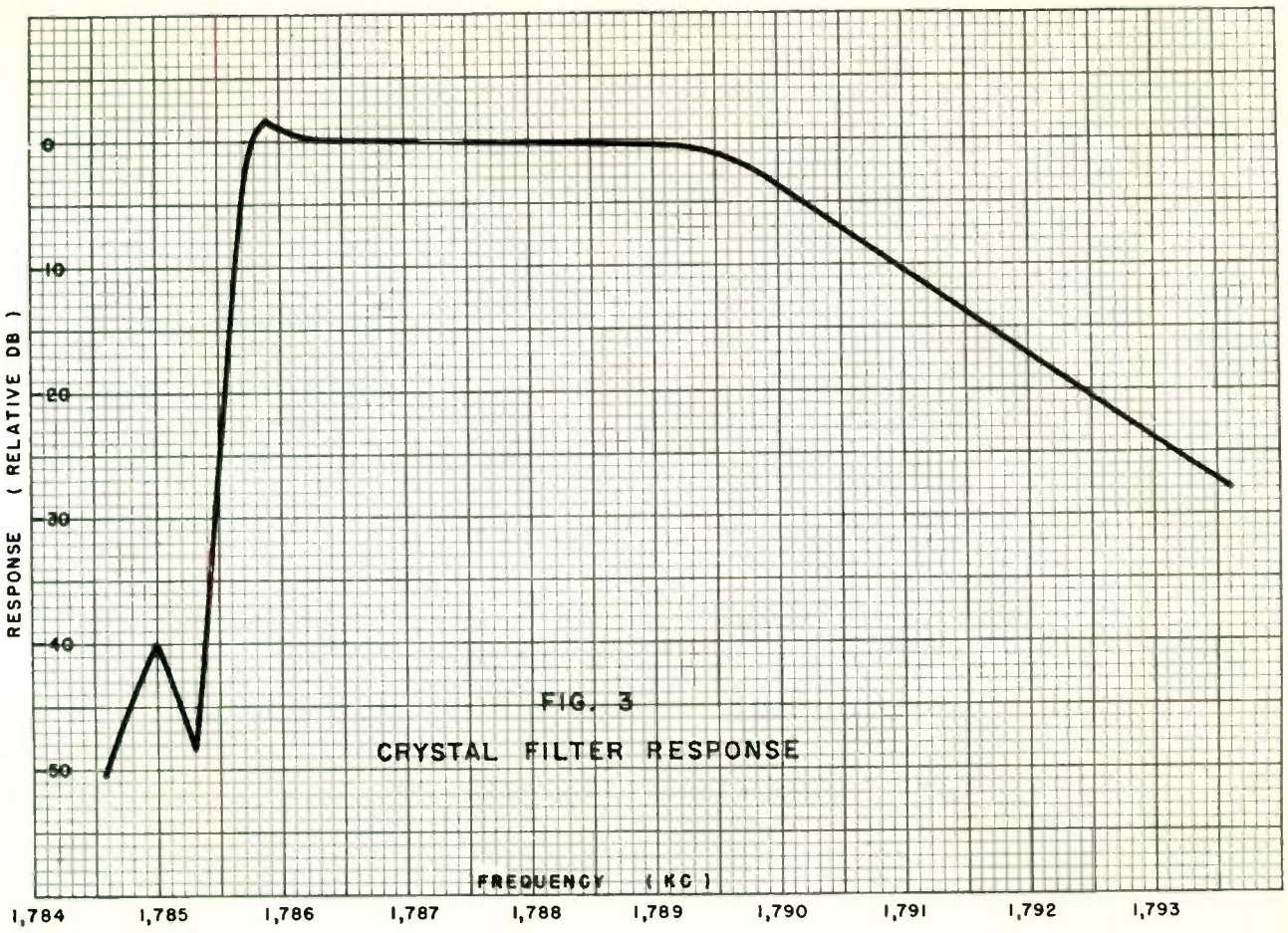


FIG. 3

CRYSTAL FILTER RESPONSE

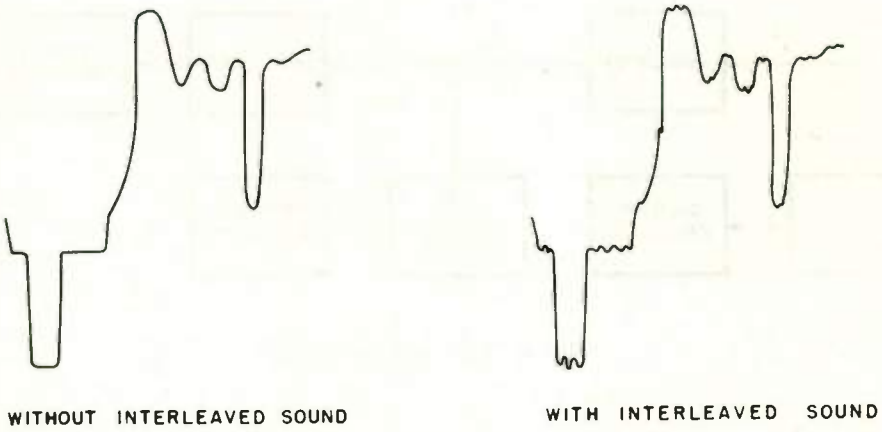


Fig. 4. Video wave forms.

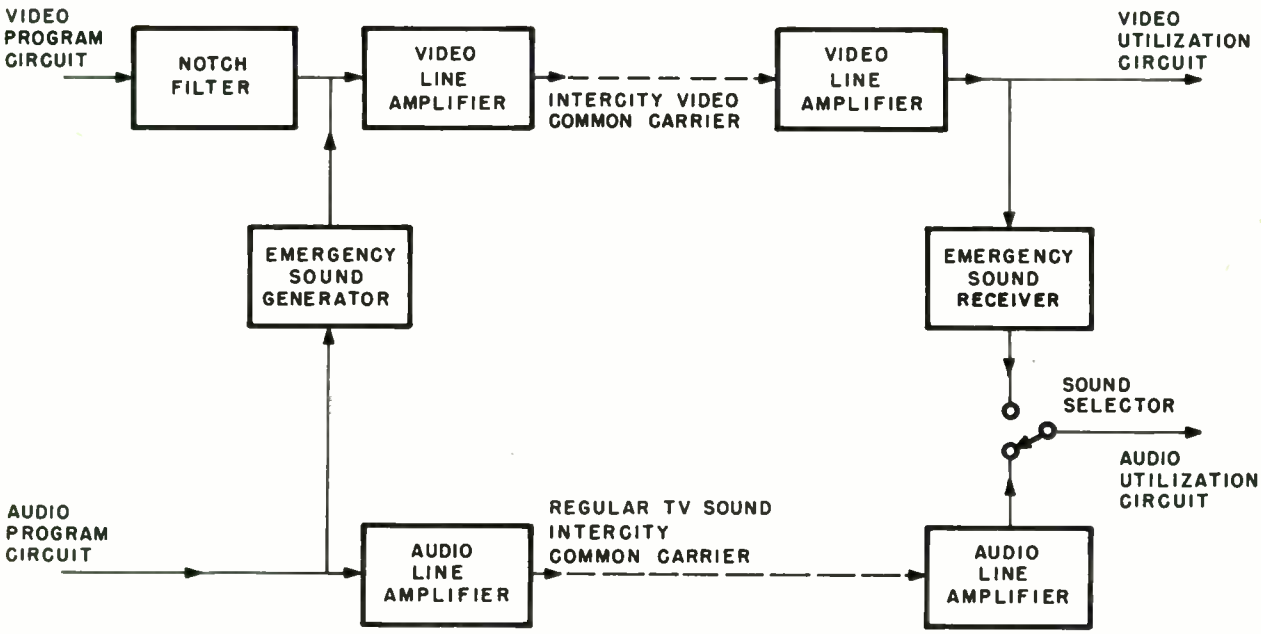


Fig. 5. Block diagram interleaved sound including notch filter.

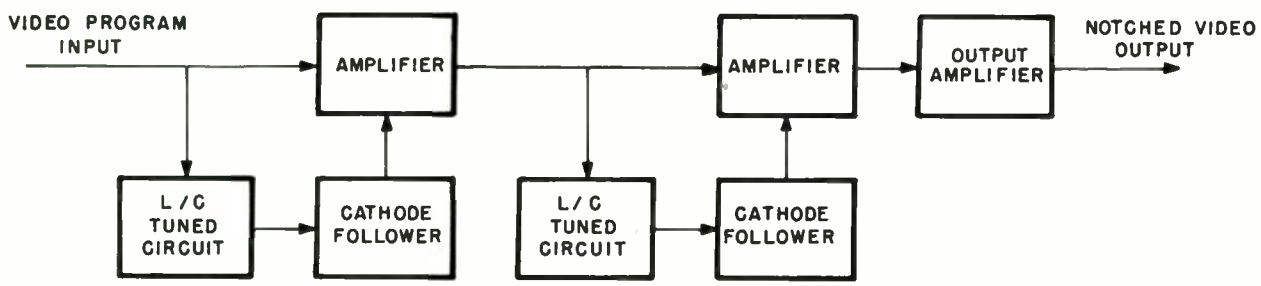


Fig. 6. Unity gain notch filter.

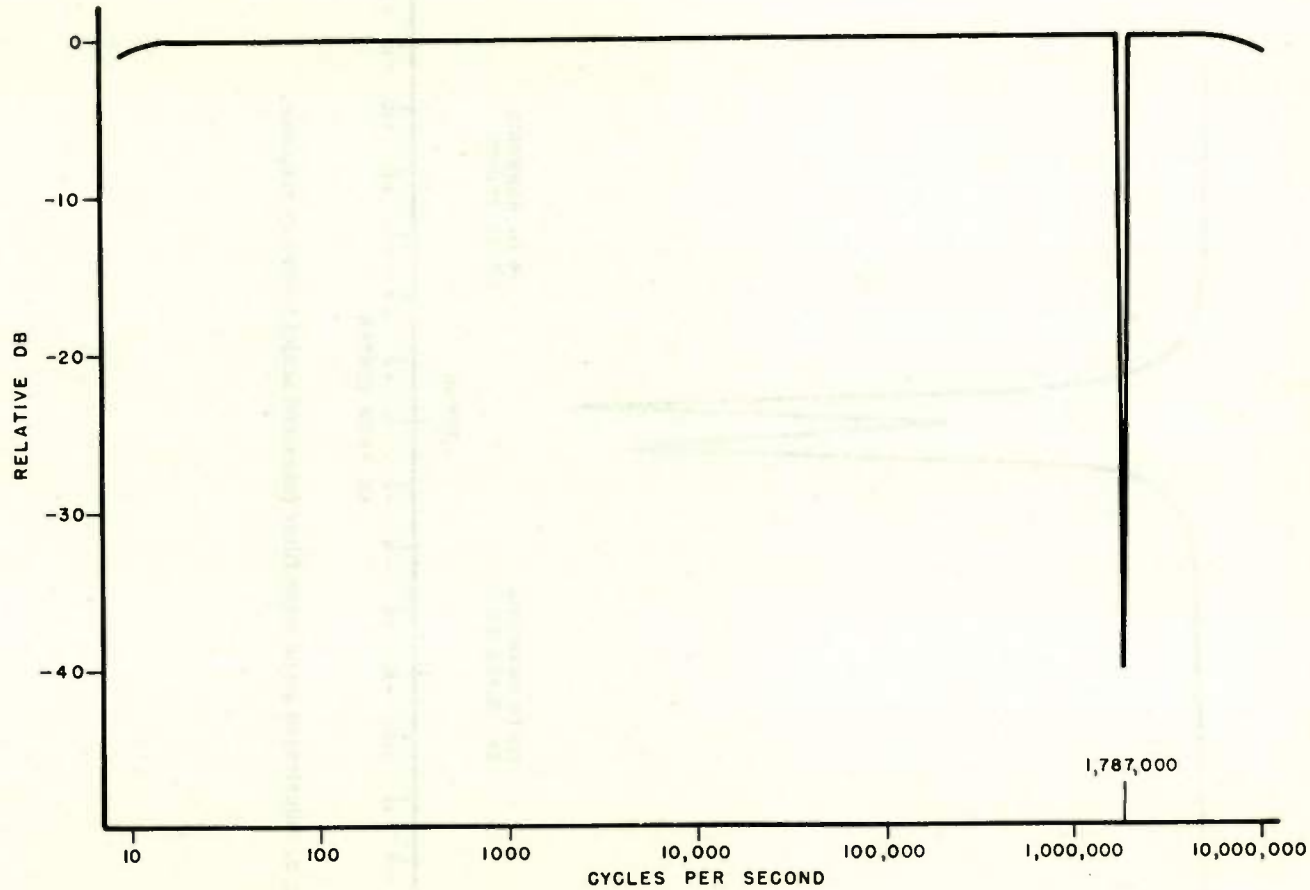


Fig. 7. Interleaved sound notch filter frequency response.

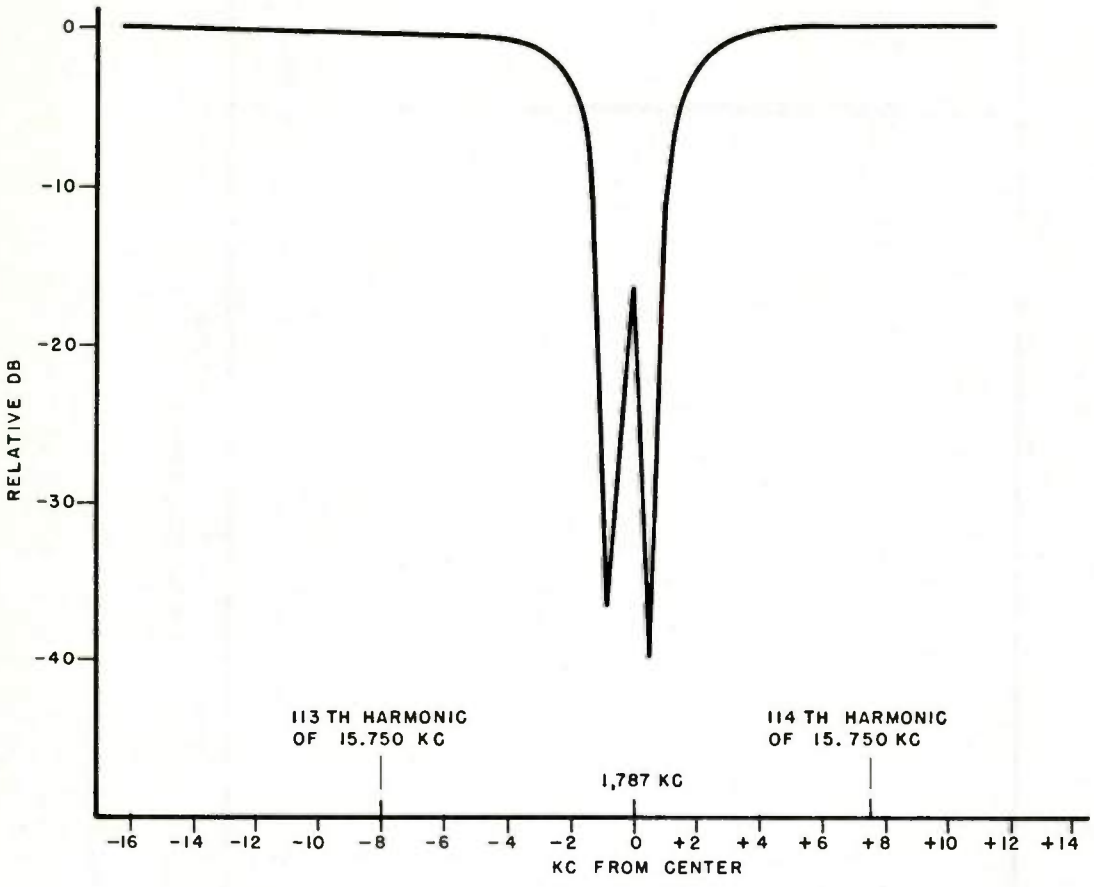


Fig. 8. Interleaved sound notch filter (expanded scale) frequency response.

A UHF TV TRANSMITTING ANTENNA FOR THE EMPIRE STATE BUILDING *

S. R. Jones+, A. Maestri+, R. W. Masters‡, and M. L. Parker‡

Introduction

Increasing pressures for more efficient spectrum utilization have led the FCC to undertake a thoroughgoing re-evaluation of the suitability of the UHF band for television transmission in highly developed areas as compared with the VHF bands.

In the planning phases of the program it was felt by many that adequate testing could be done inexpensively at relatively low power at any one of a number of locations to obtain sound engineering results from which conclusions could be drawn. Others felt that the issue was important enough to merit carrying out a complete set of observations on a typical installation in direct comparison with equivalent commercial VHF transmissions all operating from substantially the same point and under the same set of rules governing the effective radiated power (ERP). There is little room for denying conclusions arrived at with care using the latter approach, but the cost is high. The whys, wherefores, and general philosophy of the program have been set forth in a previous paper[1]. It suffices here to say that the decision was made to install a high quality UHF system on the Empire State Building where several excellent VHF installations already exist for convenient comparison in the heart of the Nation's most highly developed population center. This paper describes the antenna that was built and installed for that purpose.

The basic requirements were for high gain, omnidirectional azimuth coverage, approximately cosecant-θ elevation voltage beam contour, and high power-handling capability. It was further stipulated that the antenna be installed on the existing tower in the presence of existing antennas without creating objectionable electrical or mechanical interference, or overloading the tower, all within a time span of seven months from contract starting date to "on-air, full-power" date. These objectives were accomplished without exception, and the system has been operating since 27 November 1961, when Chairman Newton N. Minow of the FCC pressed the button officially throwing the transmitter on the air with a scheduled program.

Design Considerations

Only the larger diameter sections of the Empire State Building antenna tower were feasible locations for the addition of a high-gain UHF antenna in view of mechanical wind-load and electrical interference considerations. Of the two larger sections, the lower was preferred for aesthetic and accessibility reasons as well as to minimize the mechanical loading on the building and the probability of electrical interference. The tower sides are approximately nine feet across in this section, which makes them greater than five wavelengths at the center frequency, 575 mc, of UHF Channel 31. Those experienced in such matters will immediately appreciate the problems of accomplishing a circular azimuthal pattern with contoured vertical beam from so large a structure in the presence of the existing VHF antennas which have appreciable silhouette.

It is possible to accomplish a high degree of circularity of azimuthal pattern with only four identical radiating sources of the proper kind arrayed about a large smooth opaque square structure. The technique utilizes directive elements having their patterns oriented uniformly in a direction substantially tangential to the circle drawn through their effective radiation centers. This minimizes the space-phase, or diffraction, effects between adjacent elements as a function of azimuth angle, thus making for minimum scalloping in the resultant pattern. With the elements located on the corners of the structure

* The work reported in this paper was sponsored by the Federal Communications Commission under contract RC-9828 with Melpar, Inc., Falls Church, Va.

+ Aero Geo Astro, Inc., Alexandria, Va.; formerly with Melpar, Inc.

‡ Melpar, Inc.

the optimum beam direction of the element is approximately 45° away from the tower face, depending upon the details of its individual pattern.

The feasibility of this approach was explored initially by numerical calculations using the simplified hypothetical model illustrated in Figure 1, in which the existing structure is considered to be transparent. The four elements are coincident with the tower diagonals and displaced radially far enough to permit mounting. The resultant pattern in the quadrant where elements 1 and 2 both contribute is given by

$$E_t = E_1(\phi) + E_2(\phi) \exp j\left[\delta + \frac{2\pi d}{\lambda} \cos \phi,\right] \quad (1)$$

where $E_1(\phi)$ and $E_2(\phi)$ are the identical patterns of elements 1 and 2 modified to account for the difference in orientation, δ is the relative phase of element 2 with respect to element 1 as a zero reference, d is the separation between the two elements, and ϕ is azimuthal angle measured from the line of centers drawn from element 1 through element 2. Using $d = 5.75\lambda$, $\delta = 90°$, and representing $E_{1,2}$ by the squared cosine of angular deviation from a bearing of 45° from the respective tower faces, the pattern shown in the Figure results. The remaining three quadrants are the same as the first. Note that the total variation in the pattern is only $\pm$ 1.5 db from a circle. One of the restrictions on the scheme is that the phase progression of the element excitation around the tower must either be zero or whole multiples of 2π radians. This hypothetical model is perhaps overly idealized, but the indications obtained from it led to an experimental investigation, upon the results of which a full scale development was undertaken.

In addition to an azimuthal omni-directional characteristic, it was also desired that the antenna provide a constant field intensity of 110 db above one micro-volt per meter over a radial area from 3/4 mile out to six miles. The required transmitting antenna gain to accomplish this was calculated as a function of depression angle from the FCC F(50/50) propagation curves for an input power of 50 kw and a height of 1300 feet. As expected, this closely approximated the squared cosecant of the depression angle. The necessary power gain at a depression angle of 2.35 degrees (corresponding to a distance of six miles) is 10.5 db above isotropic. These considerations implied a shaped beam in the elevation plane with a half-power beamwidth of about 4 degrees and a pattern directivity of about 15.5 db above isotropic. This provides for transmission system losses of the order of 1 to 2 db. A downward tilt of this beam of approximately 2 degrees seemed to provide the optimum approximation.

There is no fundamental difficulty in determining an aperture distribution which will produce a pattern satisfying the above requirements. The problem is to find one which in practice lends itself readily to physical realization. Several methods of direct beam synthesis are known. Some yield an infinite aperture solution which is truncated to fit the available aperture in arriving at the approximation. Others begin with the available aperture and deduce by a curve-fitting process the distribution corresponding to a pattern that is fully specified both in phase and amplitude. Although it is impossible for these methods to take into account the engineering side of the problem with full regard for the details of practicability, they are often very helpful in pointing the way to a satisfactory solution. Woodward's curve-fitting method[2] in the present case, for example, yields a sharply center-peaked amplitude distribution of even symmetry with an associated near-linear phase distribution of odd symmetry. The result suggests a co-linear array of two uniformly-attenuated traveling-wave line-source antennas fed from a common point at the center of the aperture with the phase propagation constants of each properly adjusted to point the individual beams in the desired near-broadside direction. It has been noted[3,4], however, that some kinds of traveling wave antennas produce beams approximating a cosecant function without arraying. More recently, one of the authors has proposed and investigated such a traveling wave antenna for VHF television transmitting[5], and his results have been successfully reduced to practice in a number of cases.

Because of its extreme physical simplicity and ability to fulfill all the electrical requirements, the traveling wave approach was selected. Null fill and gain are controlled in these designs by adjusting the rate of amplitude attenuation along the array, and the beam angle is dictated by the linear phase taper imposed along the array. The amplitude distribution sought is exponential, hence, all the radiators along a given feed line are alike. The spacing between these elements then determines the beam angle.

It has been shown[6] that the directive gain relative to an isotrope of a long uniformly-attenuated traveling-wave line-source antenna having an omnidirectional broadside beam and a large over-all attenuation is given fairly accurately by

$$D_o \doteq \frac{4}{\alpha} \qquad (2)$$

where α is the attenuation of the amplitude distribution in nepers per free-space wavelength. From ordinary transmission line theory, assuming negligible series resistance, it is recalled that to a good approximation

$$\alpha \doteq \frac{G}{2Y_o} , \qquad (3)$$

where G is the conductance per wavelength, and Y_o is the characteristic admittance. The beam direction, $\hat{\theta}$, measured from the antenna axis, is given by

$$\text{Cos } \hat{\theta} = \frac{\beta_{eff}}{\beta_o} , \qquad (4)$$

where β_{eff} is the effective phase propagation constant of the aperture excitation, and β_o is the free space propagation constant. For beam directions very near broadside, (4) can be written

$$\hat{\theta} \doteq \frac{\lambda_o}{2S} - \frac{\lambda_o}{\lambda'_g} \text{ radians,} \qquad (5)$$

where λ_o is the free space wavelength, λ'_g is the effective wavelength in the loaded line, and S is the center to center spacing in free space wavelengths between successive radiators with phases alternately reversed. These formulae are all that are needed to specify a preliminary nominal design.

It was decided to use a waveguide having probe-excited resonant slots cut along the centerline of the broad face as the basic directive element for the azimuthal array and flares on the sides of the guide to shape the elementary azimuth pattern. The resulting trough-shaped aperture could be covered with a protective window. A narrow size of waveguide was chosen so that the internal electrical spacing of the slots could be just short of a half guide-wavelength, while physically maintaining a real spacing of approximately one free-space wavelength. It is necessary that the internal electrical spacing differ from an integral number of half-wavelengths in order to realize an attenuation. The slots are then reverse-coupled to bring them into a slow effective phase progression to give the desired small beam tilt. A slightly backward beam tilt was desired so the waveguide could be fed from the bottom end.

The simplicity of the design now becomes apparent, since the four guides can be fed from a single 4-way power divider via one branch line each. The real differences in the branch line electrical lengths need be no greater than 90 degrees, inasmuch as a 180-degree difference can be accomplished by putting the probes on the proper sides of the slots. Some latitude exists in selecting S and λ_o/λ'_g, but it should be mentioned that grating factor considerations establish a practical upper limit on the spacing.

Assuming that good design principles are observed, a uniform broadside, omni-directional traveling wave structure having an over-all attenuation of approximately 22 db, and a length of 24 wavelengths, will have a pattern directivity of approximately $D_o \doteq 4/0.1058 = 37.8$, or about 15.8 db above isotropic, which is the desired value. All but a very small amount of the power is radiated in the incident wave. What remains at the remote end of the antenna is reflected back and radiated harmlessly in a weak slightly up-tilted beam. With slots spaced approximately one wavelength apart, a normalized coupled slot conductance of $G/Y_o \doteq 0.21$ is indicated by equation (3).

Experimental Results

Although it was fully expected that the presence of the UHF antenna would have negligible effect upon the VHF antenna characteristics, and that the VHF radiators would not deteriorate the UHF antenna pattern unduly, it was necessary to demonstrate these things experimentally before the development could proceed. To do so, two scale models were built. One was a functional two-bay replica of the tower and VHF array at one-ninth scale, and the other was an accurate three-bay replica of the tower and VHF radiator geometries at approximately one-quarter scale. The smaller model was used for testing the operational characteristics of the VHF antenna with a mock UHF array in place, and the larger one was for working out the design details of the UHF antenna with a reasonable model of the VHF antenna in place.

An extensive pattern investigation was then carried out with the small model to determine the possible effects upon the VHF array. Patterns were measured in conical cuts about the array axis at various depression angles with the VHF model elements connected in a number of different ways, some productive of null regions which should have been quite sensitive to the perturbing effects of the UHF geometry. Some slight changes were observed in the more sensitive pattern combinations, but none of them were of large magnitude.

The pattern of the progressive-phase connection scheme actually used in practice was especially insensitive to the presence of the UHF antenna. Crosstalk measurements between the North-South and East-West sets of VHF elements and VSWR checks on the sets themselves were made to determine near-field perturbation effects of the proposed asymmetrical UHF antennas. Again no significant effects were observed.

Having thus found much encouragement, the development of the UHF antenna was started at a scale frequency using the one-quarter scale model. The work was divided into three phases, namely azimuth pattern, elevation pattern, and feed-line harness.

Azimuth Pattern

The scale factor of 1/4 was chosen for the UHF model work because of the convenience of the frequency (2300 mc) and the availability of S-Band waveguide and test equipment. A photograph of the completed model in the final stages of development is shown in Figure 2. All scale model measurements were taken with the VHF dipoles in place in order to simulate the conditions of the problem as well as possible. These were left open-circuited at their terminals since it was felt that they would respond like electrically large scatterers rather than like properly terminated receiving elements, regardless of termination.

The first phase of the development program concerned itself with the matter of evolving an acceptable configuration for the individual element which in the azimuthal array would most nearly produce a resultant circular pattern. The antenna consisted of four slotted waveguide arrays, one mounted on each corner of the tower. These were built full length in order to include realistically the scattering effects of the periodically distributed VHF elements. Three layers of the VHF antenna were deemed sufficient to represent the situation. Based on the preliminary analysis of the UHF array requirements, it was thought that the element pattern would have to be approximately 55° wide at the half-power points, with the maximum aimed at an angle of 45° from the side of the tower. Flares were utilized at the sides of the waveguide as the control parameters for shaping the radiation pattern of the individual elements.

It developed that the hypothetical model assumed for preliminary consideration was, indeed, overly simplified. Nevertheless, the final configuration which produced the nearest-circular azimuth pattern was not greatly different, and its performance was close to the predictions. It was found necessary to mount the arrays each just around the corner from the one to the rear in order to avoid back-scattering by the large discontinuity which would otherwise exist in the line of fire at the edges of the screened tower sides. Due to reflection from the face of the tower, it was found necessary to aim the array physically at an angle of 35° from the tower face and to flare the guides asymmetrically to obtain a radiation beam of the necessary shape pointing at the required 45°. The azimuth pattern of a single vertical array element is shown in Figure 3, and the total azimuth pattern taken on the nose of the beam is shown in Figure 4. It will be noted that the individual elements have a significant amount of radiation through the tower and that the pattern is rippled. The rippling effect is a matter of scattering from the VHF elements and from members of the tower structure inside the screens. The screens, being very open, are not good reflectors at the UHF frequencies.

It was learned during this phase of the program that the circularity of the pattern could be improved considerably by improving the reflectivity of the VHF screens for UHF waves. This does not seem to alter the element pattern in the main, but it reduces the ripple in it and eliminates miscellaneous transmission through the tower which seems to have an adverse effect on the over-all pattern. The screening modification was not carried out on the full scale version because of cost, possible interference with the co-licensees, and the additional wind-loading it would have contributed to the tower.

Elevation Pattern

The design of the probe-fed, longitudinal, centerline slots was accomplished using standard RG48/U waveguide. The RG48 guide is approximately one-fourth the size of the available full-scale, WR 1150 waveguide, which satisfied the size requirements for internal and external electrical spacing. The precise scale frequency was chosen so that the ratio of the free-space wavelength to guide-wavelength would be same as that of the WR 1150 guide at the center frequency of channel 31. This occurred in the RG48 waveguide at 2325 mc, thus fixing a scaling factor of 1/4.05. The scaled channel frequency limits are then 2313 mc and 2337 mc.

Working out the detail dimensions of the slot itself was a matter of patience in the laboratory. A considerable amount of

experimental design information had to be accumulated on the admittance function of the slot, normalized to the characteristic admittance of the particular waveguide, in terms of slot length, slot width, probe length, probe diameter, and probe displacement from center. The specifications set down at the end of Section 2 as an example of the use of the design equations were those actually chosen for the final design. Since the slots were to be spaced as nearly as possible to one wavelength apart, the normalized conductance value of 0.21 per wavelength became the required normalized conductance of the slot. The one-wavelength spacing makes optimum use of the aperture without overcrowding, and the 24-slot array then fits into the three spaces between four adjacent elements of the VHF array. This arrangement puts the least number of VHF elements directly in the way of the UHF antenna, thus minimizing scattering.

A preliminary array was built on the basis of equation (5) to aim the beam at a depression angle of 2.3°, with the antenna fed from the bottom end, assuming a purely conductive slot admittance and zero perturbation of the guide wavelength by the loading. An inductive post across the guide located opposite the probe at each slot was included as a convenient means for tuning the coupled slot admittance to a pure conductance at the design frequency.

Patterns measured with the array terminated in a matched load were in good agreement with the specified FCC 50/50 curve. It was found that the posts had little effect other than to control the beam direction somewhat by varying the net susceptive loading, which affects the phase propagation constant of the system. They were eliminated for purposes of simplification, and the slot spacing adjusted in a second model to aim the beam in the desired direction without the posts. It was also found preferable not to terminate the array in an absorbing load. The secondary pattern generated by the reflected wave at the remote end of the guide can be used advantageously in the narrow UHF band to improve the over-all pattern by adjusting its phase. This is done by means of a movable short at the end of the guide.

The design of the full-scale array was made exactly proportional to the scale model except for a minor adjustment of the slot length and probe depth that was necessary in order to duplicate the admittance characteristics of the scale model slot. After fabrication, an array was assembled and patterns were taken with the array terminated by an adjustable short. Figure 5 shows the full scale array mounted on a rotator for measurement of the beam pattern. The pattern at first did not duplicate that of the model, possibly because of a slight difference in inherent waveguide attenuation and possibly because of mechanical tolerances. The suitability of the probe-coupled slot radiator for the application is well illustrated in this case. The first-order effect of varying the probe depth is to change the value of the conductance. It was possible to bring the beam characteristics into good agreement with those of the model by adjusting the probe depth and the position of the short at the end of the array, and obtain a pattern which satisfied the specifications. The final pattern is shown in Figure 6.

With the full-scale array design now finalized, the other three arrays were assembled, permanent shorting sections of waveguide were fabricated for the terminations, and the probes were locked at the proper depth by jam nuts. In order to assure that the arrays were identical without measuring the patterns, the waveguide fields were probed at each of the 24 slots. Curves plotted of the results were in excellent agreement, and no further checks were considered necessary.

A final step before shipment and installation was an impedance measurement on each of the arrays at the input of the coaxial-to-waveguide transition. These were essentially identical with a VSWR of 1.8:1. By adding a matching post at a calculated position between the input and the first slot it was possible to bring the VSWR to less than 1.2:1 over the band for all arrays. The final matching at installation was left for the slug tuners provided in the feed harness described below.

Feed-Line Harness

The branch feed-line harness is very elementary. It consists simply of four 3 1/8-inch, 50-ohm coaxial transmission lines to each guide input from a common four-way power divider. The power divider has a 6 1/8-inch, 75-ohm coaxial input, with impedance transformers to accomplish matching from the junction of the four lines. The 6 1/8-inch line flares to a 9 1/8-inch line just inside the building, and progresses some 400 feet to the transmitter room. Each of the branch lines is equipped with a two-slug tuner to trim the VSWR to within 1.1:1 before entering the power divider. All the branch lines are of equal length except for the phase length necessary to set up the proper relationship among the array elements. The only difficulty in the harness was to lay it out so that it would be physically compatible with the existing geometry of the site. Measurements on the

branch lines of the harness showed agreement to within 2° of the specified lengths.

Installation and Check-out

The installation of the antenna was relatively uneventful. The antenna sections were transported in the Empire State Building elevators to the uppermost level where they were delivered to a location on the circular ice-shield platform through a port cut through the side of the conical cap for the purpose. They were then hoisted into position by ropes and bolted in place. It was necessary to do all the rigging during the off-air time of the co-licensee which often limited the available working time to two hours in the very early morning. Figure 7 shows how the arrays look in the final installation, and Figure 8 pictures the entire tower from a distance. The antenna is barely visible to the eye from street level.

Crosstalk measurements were required to be made between the new antenna and existing ones for every interested licensee on the tower. The predictions that no objectionable levels would occur were supported. The maximum level that was measured was of the order of 50 db down. Measurements made on the antenna system of the co-licensee showed no objectionable interference either with respect to VSWR or crosstalk between his own sets of elements.

Acknowledgements

Space limitations prohibit proper acknowledgement of all the many participants who contributed to the success of the work, but the authors cannot close without mentioning the tolerance and cooperation of the tower licensees and the Empire State Building Corporation during the trying period of the installation and checkout of the antenna.

The services of the Primary Committee in the person of Dr. F. G. Kear, acting officially to protect the interests of the licensees and unofficially as a scientific advisor throughout the program, were indispensible. The devotion of the various technical and support personnel within Melpar, many of whom worked long overtime hours, was a vital ingredient in accomplishing the job. The installation was supervised by C. L. Padgett, who did much of the mechanical design, and the program was coordinated by H. C. Wilson.

References

1. A. G. Skrivseth, "New York UHF-TV Project of the FCC' Federal Communications Commission, 11th Annual Broadcast Symposium, IRE Transactions on Broadcasting, December 1961.

2. P. M. Woodward, "A Method for Calculating the Field over a Plane Aperture Required to Produce a Given Polar Diagram", Journal I.E.E., 93, Part III-A, No. 10, March-May 1946, pp. 1554-1558.

3. G. C. Southworth, "Principles and Applications of Waveguide Transmission", D. Van Nostrand, 1950, p. 427.

4. A. S. Dunbar, "On the Theory of Antenna Beam Shaping", Journal Applied Physics, Volume 23, No. 8, August 1952, p. 847.

5. R. W. Masters, "Traveling Wave Antenna", U. S. Patent No. 2,947,988, 2 August 1960.

6. R. W. Masters, "Traveling Wave Antennas with Application in Television Transmitting", Doctoral Dissertation, University of Pennsylvania, 1957.

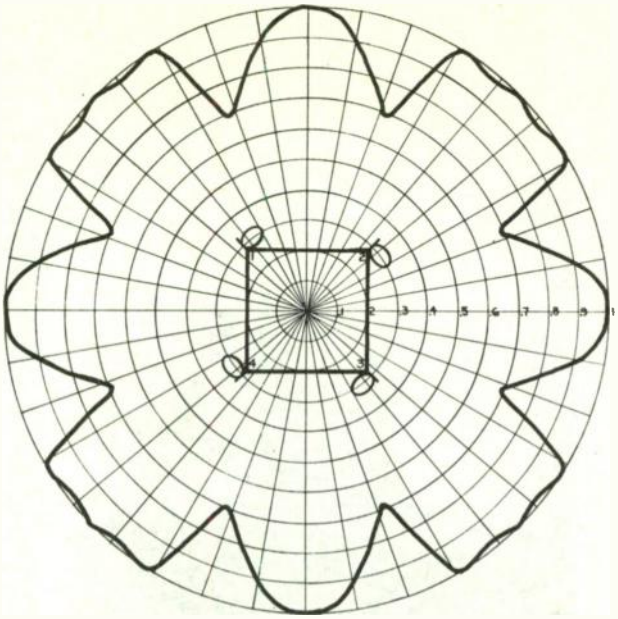


Fig. 1. Hypothetical array and resultant azimuth voltage pattern.

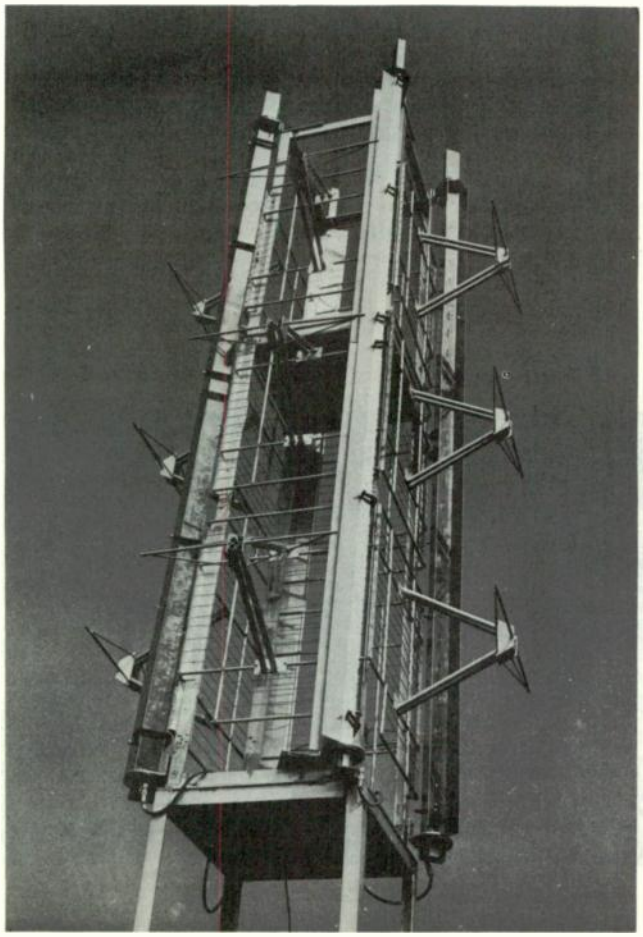


Fig. 2. Experimental tower model built to 1/4.05 scale.

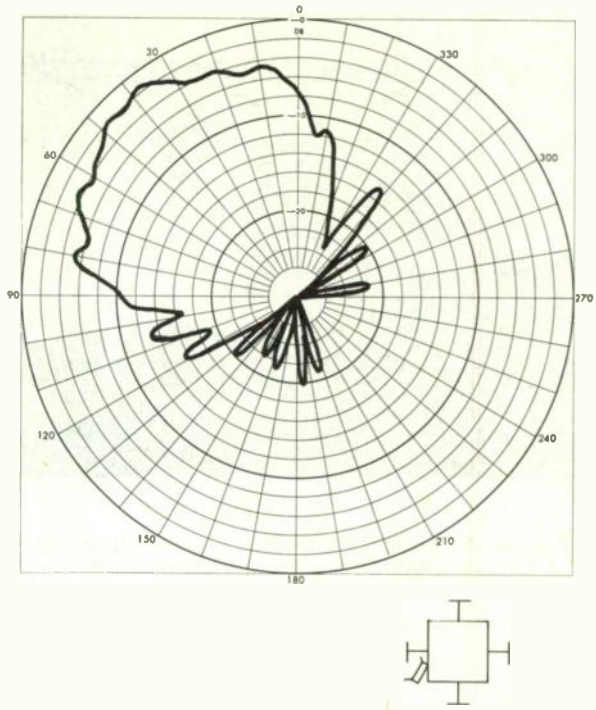


Fig. 3. Typical single-element azimuth pattern (db) measured on beam nose of scale tower model.

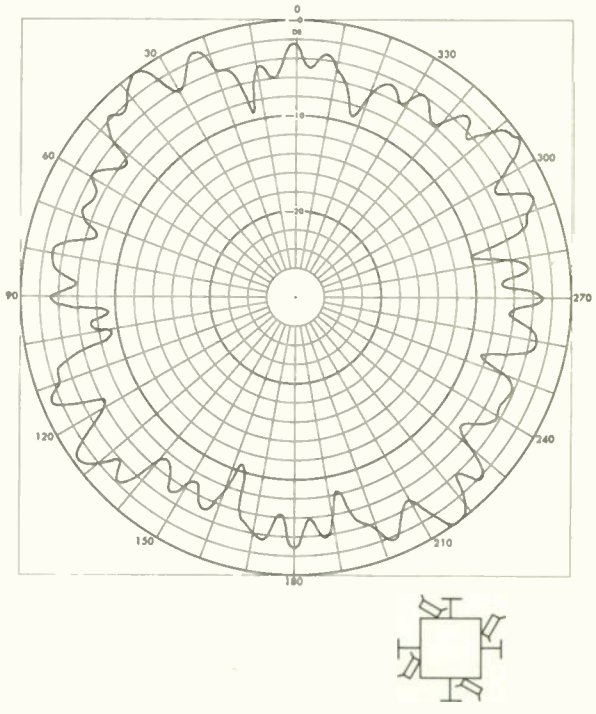


Fig. 4. Typical resultant azimuth pattern (db) measured on beam nose of scale tower model.

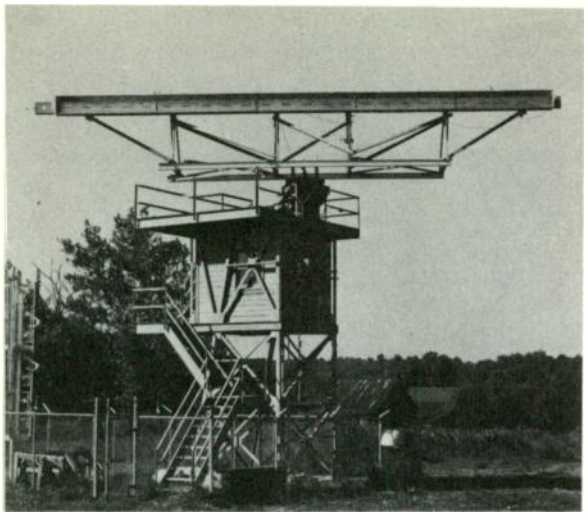


Fig. 5. Single element of full-scale array mounted for elevation beam pattern measurement.

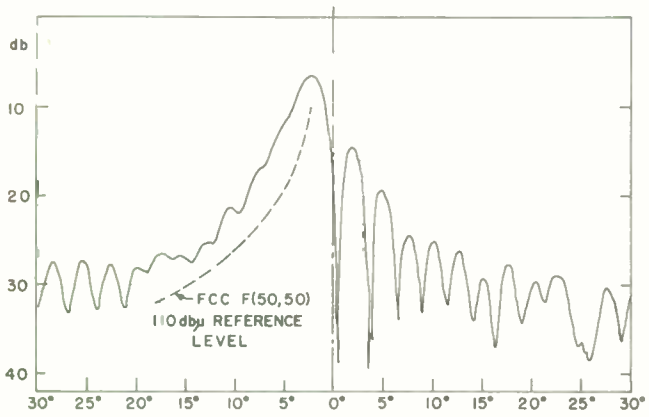


Fig. 6. Elevation pattern (db) of full-scale antenna element at 575 mc.

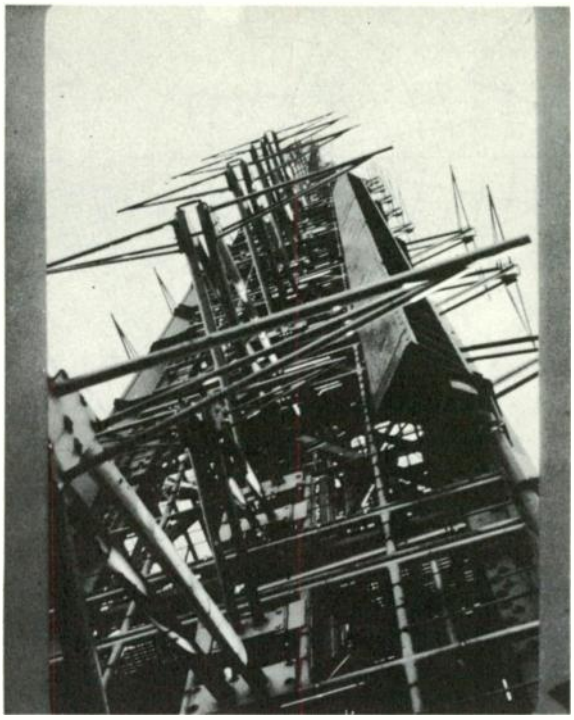


Fig. 7. Close-up view of full-scale antenna element installed on Empire State Tower.

Fig. 8. Aerial view of Empire State Tower antenna system.

A REVIEW OF SOME OF THE RECENT DEVELOPMENTS IN COLOR TV

Bernard D. Loughlin
Electronic Research Consultant
Huntington, New York

There appears to be a current upswing in the interest in color TV in U. S. A. This follows a period of about 5 or 6 years during which the general degree of activity in the color TV field in the U. S. A. may have appeared externally to have been low. However, some significant developments have been made during this period. To assist engineers now becoming interested in color TV, the organizer of the PGBTR session requested a review paper of some of these more recent developments.

Since this paper is for a PGBTR session, items relating specifically to receivers, including picture tubes and receiver circuitry, will be covered. Because the major items of discussion relate to specific picture tube developments (and associated circuitry), most of the material will be grouped according to the type of color display. This material will attempt to review the technical developments and to stay away from any direct economic comparisons.

System standards do have some effect upon receiver design. Therefore, for general interest, the color TV systems under serious consideration for European color TV will also be briefly reviewed.

For the benefit of those who may not have been closely associated with color TV in the recent past it has been requested that a brief review of some of the fundamentals of color TV be given first.

Theory Review

There have been many books and articles written which cover this subject and for more detailed information the reader should consult references 1 through 6, as well as the two IRE color TV issues of October 1951 and January 1954.

NTSC Color Signal

The U. S. A. color television standards call for the signal which was recommended by the industry group known as the National Television System Committee, and this signal is generally referred to as the NTSC color signal.

Luminance Signal. The NTSC color signal contains a wide-band luminance signal which is transmitted substantially according to normal monochrome TV standards, providing a compatible system. The wide-band luminance signal (Y) is produced by combining the red, green and blue primary color signals (R, G, B) in proportion to their relative contributions to luminance.

Color-difference Signals. A pair of narrower bandwidth color-difference signals are added to the wide-band luminance signal to provide the coloring information for the picture. Color-difference signals indicate how the primary color signals differ from the luminance signal and therefore signals such as R-Y, G-Y and B-Y are color-difference signals. Only two of the three color-difference signals just cited can be considered as being independent of each other, and the third color-difference signal can always be derived from the other two. Therefore, it is only necessary to transmit two independent variables to represent the color-difference information. The NTSC signal can be conveniently considered as transmitting R-Y and B-Y color-difference signals by modulating them upon quadature components of a sub-carrier. Thus the basic equation for the NTSC signal is substantially as given in Figure 1.

Scale factors. Since, the three primary color signals R, G and B are three independent quantities, a scale factor interrelating the three quantities must be established before equations involving their combination have significance. It is an accepted standard in both the fields of colorimetry and color TV that the implicit scale factors are such that the primary color signals equal each other for a color corresponding to the reference white of the system (CIE Illuminant C for NTSC standards). Further by convention the coefficients of the luminance signal are normallized to add up to unity - so on reference white: R=G=B=Y.

Encoding for compatibility. To insure maximum compatibility the color-difference signals R-Y and B-Y are modulated upon the same sub-carrier frequency - specifically upon components in quadature. The sub-carrier frequency is chosen to be quite high so its

resulting dot pattern is of fine structure. In addition, it is chosen to be a specific frequency (an odd multiple of one-half horizontal scanning frequency) which produces interleaving of the dots between successive lines and successive frames. This provides a minimum of visibility in a channel reproducing the luminance signal (such as a monochrome receiver). This technique of interleaving for low visibility is similar to the technique of offset carrier employed to minimize interference between co-channel television stations.

For maximum compatibility the color-difference signals are modulated on the sub-carrier in a manner which can be called suppressed-carrier modulation because the sub-carrier disappears when the color difference signals are zero. This condition exists when the color to be reproduced is reference white. Correspondingly this suppressed-carrier operation results in a low amplitude of sub-carrier signal for the more frequent pastel colors which are not highly saturated.

By using the above features, the modulated sub-carrier signal can be transmitted in the upper portion of the luminance signal bandwidth. Thus it can be said that the NTSC signal is a simultaneous band-shared color television signal using a wide-band luminance component and narrower bandwidth color-difference components transmitted as suppressed-carrier modulation on a pair of low visibility quadature sub-carriers.

Chrominance Signal. The composite modulated sub-carrier signal is also called the chrominance signal which can be represented vectorially as shown in Figure 2. Besides thinking of this as a pair of quadature modulated components it is also convenient, for some applications, to realize that the chrominance signal has a resultant phase which is related to the hue of the color to be reproduced and an amplitude (with reference to the luminance signal) which is related to the saturation of the color.

Other features. There are three other items that should be noted about the NTSC signal. In order to correctly decode the chrominance signal it is necessary to have a phase reference and this is provided by a short burst of the color sub-carrier at a reference phase of -(B-Y) which is transmitted during the back porch.

A second item to be noted is that the gamma correction in use involves individual pre-correction of the red, green and blue signals to compensate for the normal non-linear characteristic of electron guns. This type of gamma correction permits linear decoding in the receiver to give red, green and blue signals which are directly suitable for obtaining correct colorimetry over the complete color gamut on a proper 3 gun display. It also permits linear processing to give such "ideal colorimetry" on 1 gun displays with narrow-angle sampling. While it is not always appreciated, this is the only point in the signal specifications that seems to favor a 3 gun type of display. However, in reality the specifications are so worded that this form of gamma correction is not required, but instead, what is impliedly required is some reasonable degree of compatibility for a display needing this type of gamma correction for correct colorimetry.

A third, and somewhat more sophisticated item, that should be noted is that the NTSC chrominance signal contains a medium bandwidth component called "I," which corresponds to a chrominance signal axis which is 33 degrees from R-Y in the orange direction, and a narrow bandwidth component called "Q," which is 33 degrees from B-Y in the magenta direction. This is included so that medium sized color detail, along the chrominance axis producing maximum visual chromaticity detail, can be reproduced (in proper design receivers) without color crosstalk. However, current receiver designs do not make significant use of this feature. Instead the signal is generally treated as an equal-band chrominance signal having a bandwidth somewhat between the narrow Q value and the medium I value.

Receiver Decoding for Three-Gun Displays

Basic Shunted-monochrome Arrangement. Decoders for 3 gun displays employ two shunt channels - one which handles the wide-band luminance signal and the other which handles the chrominance information. The wide-band luminance signal is directly applied to the 3 guns (generally to the cathodes) to produce a wide-band black and white picture. The chrominance signal is selected by a suitable bandpass amplifier and applied to a pair of synchronous detectors to derive two color-difference type signals. These derived color-difference type signals may be (although they need not necessarily be) R-Y and B-Y signals. This pair of color-difference signals are then matrixed to obtain 3 color-difference signals which are usually applied to the grids of the 3 electron guns, to add

coloring to the picture. Thus the basic block diagram is as shown in Figure 3.

Color-difference Matrix. The matrix operation to get 3 color-difference signals from the 2 detected components can take a variety of forms. If the R-Y and B-Y components are directly detected, then the matrix is merely required to properly intermix these signals to obtain the G-Y signal - and may possibly amplify the R-Y and B-Y components without intermixing. On the other hand the detected components might correspond to some other angles than those for R-Y and B-Y and these angles may be somewhat off of a quadature relation. In this case the matrix is required to intermix the components in order to derive each one of the desired R-Y, B-Y, and G-Y signals. While this may sound more complicated, it can actually be accomplished quite simply (see ref. 3, sect. 15-11).

Color Sync. For the synchronous detectors to operate correctly a source of reference sub-carrier at a specific phase is required. This is provided by the part of the decoder which is normally known as the color sync section. The color sync portion contains a gating circuit to select the burst (independent of the chrominance signal) and this gated burst is then used to control (or lock) a stable oscillator, generally through an APC (automatic phase control) loop. Such a loop operates much like the normal horizontal deflection synchronizing circuits now used in all television receivers.

At various times great concern has been expressed over the fact that such a color synchronizing circuit is required for the NTSC signal. However, extended experience in this country does not warrant such concern. The color synchronizing portion of the receiver is in general, quite stable and there is much experience which indicates that with weak signals the color synchronizing function performs substantially better than normal horizontal and vertical deflection synchronization functions.

Other features. There are three other items which should be noted which stem from the two shunt channel arrangement. First the wide-band luminance channel normally uses a delay line to provide delay equalization between this wider-bandwidth channel and the narrow-band chrominance channel.

As a second item, convenient hue and saturation controls are generally provided in the decoder. Adjustment of reproduced hue is obtained by a simple phasing control in the color synchronizing circuits. This is frequently accomplished by detuning a circuit in the burst-gate channel or in the APC phase detector. Reproduced saturation is adjusted by a simple gain control in the chrominance channel. Some receivers have included an automatic chrominance control (ACC) circuit to keep the reproduced saturation approximately constant in spite of items such as multipath and receiver detuning - which might vary the amplitude of the chrominance signal compared to the luminance signal.

As a third item, it is desirable to disable the color decoding function during monochrome reception. This is so because the color sub-carrier oscillator runs free and can result in spurious beats with high frequency luminance detail which are annoying because they are not interleaved (as they are on color reception). Therefore a color killing function is used which cuts off the chrominance channel during monochrome reception by sensing the absence of the color burst.

Receiver Decoding for One-Gun Displays

One-gun displays are generally operated in a sequential manner, so that the electron beam impinges upon the color phosphors in some repetitive sequence. Two general modes of operation have been employed and these have been classified as continuous color sequence (CCS) and reversing color sequence (RCS). In the continuous color sequence each primary color occurs only once before the sequence is repeated, such as: GRBGRB. In the reversing color sequence at least one color is repeated twice before the overall sequence repeats - for example: GRGBGRGB.

Examples of previous displays using continuous color sequence are the early RCA one-gun shadowmask tube, the Apple tube, and the Chromatron tube with third harmonic gating. An example of operation using a reversing color sequence is the Chromatron when sinusoidally color switched and operated for maximum light output (self gating) without third harmonic gating.

CCS Operation-direct application. A three color picture can be obtained by direct application of the NTSC signal to a CCS display, if the color switching is synchronized to the NTSC color sub-carrier frequency and the sequence has the right order. By using a selected amount of chrominance boost and a selected phasing, it is possible to reproduce one axis of colors (such as the orange-

to-cyan axis) with reasonable colorimetry, but errors exist for other colors (see Fig. 3 of reference 5). Besides these errors, such operation results in greater visibility of spurious signals and noise. Also on monochrome reception non-interleaved crosstalk components plus a non-interleaved 3.58mc color sampling pattern are produced (which are more visible than on color reception).

Operation similar to the above can also be practiced with the Apple tube by conversion of the chrominance signal frequency to the proper frequency and phase needed for direct decoding by the phosphors strips. This involves conversion of the color sub-carrier to a frequency around six megacycles (under control of the indexing information) for the present design of tube. Such a circuit uses shunt monochrome and chrominance channels and therefore permits effective color killing to reduce some of the above cited crosstalk components on monochrome reception. (Also the color sampling pattern is stationary.)

CCS Operation - processing for correct color. One method to get correct colorimetry would be to decode to simultaneous R, G and B signals and then sample these at the proper times needed by the display. Also, a method can be employed in which only color-difference signals are decoded and these are sampled at the proper time - the luminance signal being applied to the display by a shunt channel. While such two step methods (decoding followed by sampling) have received some consideration, methods involving direct signal processing have seemed most attractive.[5]

Direct processing involves, as one step, the addition of a selected color-difference signal to the luminance signal to produce a new monochrome signal (M) having more nearly equal weighting of the primary color signal components. As a second step, the NTSC chrominance signal can be modified to a symmetrical sub-carrier form by a circuit known as an elliptical gain amplifier. The basic block diagram for such a signal processing circuit is illustrated in Figure 4. If the color sequence occurs at some rate other than the transmitted color sub-carrier frequency of 3.58 megacycles the system may include a frequency conversion step for the chrominance signal (such as when using the Apple tube).

Strictly speaking the above linear processing when used with wide-angle sampling (such as may result with self gating by the phosphor strips or spots) does not result in exact colorimetry, particularly on saturated colors, due to non-linear (gamma) effects. This has been compensated for in a variety of ways. First, merely a 1st order correction for the wide-angle sampling can be made by increasing the chrominance channel gain to a value which provides a compromise between pastel and saturated colors. Second, the "angle of sampling" can be effectively reduced on saturated colors by using additional harmonics of the chrominance signal. Third, a CW gating signal can be used to narrow the angle of sampling - the required amplitude may be small. Fourth, black guard band can be built into the tube between phosphor strips or spots. Fifth, a non-linear correction can be applied to the monochrome signal (proportional to amplitude of chrominance signal) which has been called either saturation correction" or "diode correction." [7, 11]

Of course to make a complete receiver it is necessary to include a color synchronizing section to obtain a reference color sub-carrier signal. This reference signal is required for correct processing and also to control either the color switching or the color sub-carrier frequency conversion step.

RCS Operation. In this case direct application of the NTSC signal does not result in a three color picture. The two step decoding followed by sampling process can be used (as discussed under CCS), but again direct signal processing methods have seemed more attractive.[5]

It can be shown that such direct processing methods need (for 3 color reproduction) a pair of chrominance components at frequencies harmonically related to the basic color-switching frequency. The sequence cited previously, namely GRGBGRGB (which is obtained with sinusoidal switching of a Chromatron having green for the unswitched phosphor) requires a fundamental frequency R-B component and a second harmonic G-M component. For optimum results such chrominance components should be of fixed phase and correspond only to a selected single color-difference signal axis. Such processing is produced by a circuit called an axis selector, which is like the so-called elliptical gain amplifier in which the gain is large for the desired axis and very low for an axis in quadature.

For complete processing monochrome signal correction is also required. A block diagram for such an arrangement is shown in Figure 5. As illustrated such an arrangement also requires a color sync section as before.

Picture Tube and Related Circuit Developments

The developments reviewed here are those which have either occurred or been released during the past five or six year period. Where applicable, the previously described status of the tubes and related circuits are breifly discussed. Since the discussion will be in alphabetic order, it might be said to provide the ABC'sz of color picture tubes.

A is for Apple

The Apple tube is a beam-indexing color display using vertical color stripes so that horizontal scanning makes the beam sequentially imping upon the various color phosphors. It is usually described as having successive red, green and blue phosphor stripes repeating in a simple sequence so that CCS operation results. Black guard bands are provided between successive stripes to reduce the equivalent sampling angle. Beam-indexing information (describing the beam position) is used to control the decoder so that the proper instantaneous beam current is produced as the beam scans over the successive phosphors.

Apple Tube as described in 1956. This Apple tube used secondary-emission index stripes, as illustrated in Figure 6. One wide indexing stripe was provided per triplet of color phosphors, so the basic index information frequency was the same as the triplet frequency (being 6 or 7 mc. depending upon design of tube). To obtain the index signal clean and independent of the color picture information a pilot carrier system was employed having a frequency around 42mc. One of the sidebands produced by the indexing stripe modulation of the pilot carrier was selected, amplified and used to control the decoding by a system such as the one illustrated in Figure 7.

To prevent undue modulation of the pilot carrier signal by the color signal a two beam arrangement was employed. One beam, the writing beam, was modulated by the color signal, and the other beam, the pilot beam, was modulated by the pilot carrier (at approximately 42mc). These two beams were provided by a single electron gun having two control grids (in the same plane with apertures displaced from each other). Further, to prevent crosstalk of writing beam harmonics into the indexing channel the sideband amplifier frequency (and therefore the pilot carrier frequency) was selected to interleave between the harmonics of the writing beam chrominance signal.

Because of the phase shift versus frequency characteristic of the sideband amplifier it was necessary to maintain the scanning speed quite close to a constant value - otherwise undesirable hue shifts would result. This required tight tolerances on both linearity and size of horizontal scan. Further, it was found desirable to print the indexing stripes so that they did not align with the same part of the color triplet at all points over the face of the tube. Instead a controlled index displacement was employed which compensated for both imperfect tracking between the two beams and transit time effects.

Two other characteristics of the secondary-emission index Apple tube arrangement should be noted. First, due to the high gain 48mc. sideband amplifier there is a potential interference problem due to pickup of radiated signals (the picture tube providing an antenna). Second, both the color saturation and contrast ratio are somewhat limited due to the exitation of the phosphors by the pilot beam current.

Recently Released Developments in Apple. An alternative Apple tube has been recently described using index stripes which employ UV radiation for indexing information instead of secondary emission. This arrangement is illustrated in Figure 8. A photo multiplier tube views the rear side of the phosphor screen structure and picks up the UV radiation produced by the P16 phosphor used for indexing.

Another difference is that the indexing stripes are placed closer together than once per triplet - being put after each pair of color phosphor stripes. Thus, for a triplet frequency of 6mc. the basic indexing frequency will be 9mc. In this arrangement the average phase of the 9mc indexing signal is independent of the applied chrominance signal.

In this system there exists an ambiguity which is resolved by providing a special index structure at the start of each horizontal line scan (which may actually turn around the edge of the picture tube). This starting section provides a non-ambiguous set of indexing pulses (which may be at either a 3 or a 6 mc rate). Phase modulation of these starting pulses is prevented by blanking out the color signal during the starting period.

The UV or photo-indexing type of system provides a series of advantages. The indexing information can be used

directly at the indexing frequency without a pilot carrier system - thereby reducing the complexity. Since the index information is collected in the form of photons the interference and transit time problems are eliminated.

The sweep tolerances have been considerably eased by using a phase compensated loop which permits approximate cancellation of the static phase shift versus frequency characteristic of the indexing loop. This permits tolerances of the order of ± 5% in size and sweep rate for the horizontal scan as compared to requirements of about ± 1/2% for the previously published Apple system.

The phase compensated loop performs by developing a final chrominance signal whose frequency (and therefore phase) is a difference between an index signal and a modified (in frequency) chrominance signal. By correct design this difference between the static phase shift versus frequency characteristics of the two branches of the loop can be made to cancel each other. This type of circuit is shown in the Figure 9 block diagram of an experimental Apple receiver.

With the photo-index arrangement it is possible to use a one beam electron gun, which eliminates the tracking problem between the writing and pilot beams. However, to prevent complete loss of the indexing information it is necessary to insure a minimum average beam current at all times. In the non-integral index stripe spacing this minimum is placed only upon the luminance signal and does not prevent instantaneous beam cut-off by the chrominance signal. Therefore in the photo-index arrangement the minimum beam current requirement for indexing limits only the contrast and not the color saturation.

Improvement has been made by using elliptical electron optics in the gun to provide a spot having a very narrow horizontal dimension with a larger vertical dimension. This permits a significant increase in brightness before saturation is lost (due to spot size blow up) and has resulted in an increase in brightness in excess of 2 to 1 (compared to previous designs). Also the improved electron optics, combined with the elimination of the dual beam feature, permits the use of scanning yoke having about the same size as one employed in monochrome receivers, but perhaps requiring more precise construction.

The current Apple tubes use a conventional black and white bulb and have a 74° deflection angle. Certain steps, particularly with regard to the electron gun, have been taken towards a 90° deflection angle design and the engineers involved are convinced that the electron optics for such a design are feasible.

B is for Banana

The Banana tube color display system was developed (in England) to provide a color display requiring a simple cathode-ray tube. The cathode-ray tube merely provides a single line of color information and is viewed through an optical system providing field scan by mechanical means.

The cathode-ray tube has a long cylindrical bulb with an electron gun at one end and with the phosphor screen along one side of the cylinder, such as illustrated in Figure 10. This "banana" type construction is necessary to fit the cathode-ray tube into the middle of the rotating drum used to provide vertical scan. The mechanical layout, including the rotating drum with its three cylindrical lenses and the parabolic viewing mirror, is illustrated in Figure 11.

Due to the small angle of approach of the electron beam to the screen at the far end of the tube, an elliptical spot can result with poor resolution. To take care of this a permanent magnet deflecting field is applied to the tube, with a suitable gradiant, to increase the angle of approach of the beam to the screen near the end of the tube. This is illustrated by the trajectories shown in Figure 10.

Color switching in the Banana tube is accomplished by spot wobble perpendicular to the direction of line scan (the three color phosphors being adjacent stripes). Published data indicates that the systems demonstrated employed sinusoidal spot wobble, thereby resulting in a reversing color sequence (unless third harmonic gating is employed). It appears that the two-step system of decoding to simultaneous signals followed by sampling has been used in the laboratory to establish the ultimate performance of the Banana tube type display. Complete receivers appear to have used third harmonic gating to provide CCS operation, with direct signal processing much as illustrated by Figure 4. RCS operation with suitable signal processing has been considered, but it is not apparent from the literature whether such an arrangement has been demonstrated.

While this type of system has resulted in a simplier cathode-ray tube

it does have related mechanical and optical problems because of its mechanical scan system. Also coupled with this mechanical scan is the fact that fast decay phosphors must be used in the CRT to prevent vertical smear. This dictated the use of the new sulphide group of phosphors which produces a somewhat desaturated green.

As a summary of the Banana tube situation it is interesting to consider a direct quotation from page 585 of reference 20 which, with typical British reserve, states: "The Banana tube display system as described here has not yet outgrown the laboratory stage. Whether it will ever be applied in domestic color receivers in competition with the now well established shadow-mask tube is still very much an open question. The answer to it must depend to a large extent on a realistic assessment of its present performance and its potential advantages."

C is for Chromatron

The Chromatron is a post deflection focus (PDF) color tube in which the electric field set up between the phosphor screen and a closely spaced grid of wires focuses the electron beam upon a specific color phosphor. Because of this focus action about 80 to 85% of the total beam current strikes the screen. The Chromatron is generally described as a one-gun tube in which a potential difference applied between alternate wires is used to distort the electron lenses to produce color switching or focus upon a different color phosphor. However, three-gun varieties have also been considered in which the angle of approach determines the particular phosphor upon which the electron beam is focused. During the period being reviewed here effort has been almost solely on the one-gun version using sinusoidal color switching at a 3.6mc rate. However, I have been informed that now (1962) 3-gun Chromatrons are again being produced.

Older type 1-gun Chromatron. In these older tubes the phosphors were deposited (by conventional screen printing techniques) in paralled stripes on a flat glass plate. A suitable frame held the parallel closely spaced wires in a flat plane just behind this phosphor screen. This structure was then mounted in a 70°-22 inch metal cone tube resulting in a layout such as illustrated by Figure 12. The resulting picture was only about 12 x 16 inches since the rectangular color structure had to be fitted inside the round metal cone.

While several variations of the tube were made, the one generally considered had horizontal stripes with green as the center or unswitched phosphor color. The wire grid spacing used (40 to the inch) resulted in approximately 450 green stripes in the picture height and only about 225 red and 225 blue stripes. This gave a rather coarse line structure particularly on saturated reds (such as lips).

The resolution and brightness of the tube was limited since the electron gun was basically a low voltage gun (approximately 5kv). Further, the contrast of the image was limited due to multiple reflections in the several glass plates, plus the effect of both secondary electrons from the wire structure and reflected primaries from the screen structure.

Due to the rather large capacitance between the grid wires sinusoidal color switching at a 3.6mc rate results in a quite large circulating current. This required rather significant switching power, particularly because of the loss in the connectors used to couple the high capacitance wire grid inside of the tube to the resonating inductive element outside of the tube.

A variety of decoding circuits were used with the previous Chromatron. These included the two-step decoding plus sampling, CCS operation with third harmonic gating and RCS operation using circuits of the general arrangement shown in Figure 5. With the direct processing circuits it was necessary to compensate for the color balance on white due to the differing phosphor efficiencies. This was accomplished either by a color filter in front of the picture tube or by adding a CW signal to the video signal, or by a combination of these.

1961 1-gun Chromatron. In the type tube shown in 1961 the phosphor stripes were vertical and deposited directly on the face plate of the tube. The closely spaced vertical wires were mounted on a cylindrical frame just behind the face plate. The face plate shape was a section of a torus having a larger radius of curvature in the vertical direction than in the horizontal direction, and these curvatures cooperated with the cylindrical grid to obtain an approximately uniform deflection sensitivity over the screen. The general structure is illustrated by Figure 13.

The phosphors were deposited on the face plate by a technique involving

electronic printing (known as the married parts technique). This permits the exact electron optics of a particular wire grid to determine the phosphor location for the tube using that particular wire grid and results in proper phosphor stripe variations over the screen, thereby producing a very significant improvement in color uniformity.

This 1961 version of the tube used a 90°-22 inch glass rectangular bulb with only about 1/2 inch loss in picture area around the edge. This bulb differed from the conventional black and white tube bulb predominantly by the required face plate curvatures. The wire pitch was increased to 56 per inch and this, with the larger horizontal dimension, resulted in about 1000 wires across the picture. The phosphor structure used had red in the unswitched position giving about 1000 red stripes, with about 500 green and 500 blue stripes, across the picture. Further, the wire frame structure was simplified, particularly by using a frit to cement the wires to the frame and then using a combing technique to separate every other wire (a mass production weaving technique).

The color switching power was reduced by mounting the resonating inductor inside of the tube. A pair of air core toroid coils were used, one mounted above the wire frame and the other below. The resulting circulating current passing through the connectors to the external circuit was thereby reduced to a small fraction of the total circulating current. The power required for color switching was produced with about 30 watts input to the power amplifier, in spite of the increased capacitance with the finer wire spacing, and the use of 25kv.

The resolution and brightness of the tube was increased (at least 2½ to 1 in brightness for the same resolution) by using a bi-potential type of operation. The electron gun operated at 25kv, together with the aquadag surface of the tube and a hat shield placed just before the wire grid frame (such as illustrated by Figure 13). The required horizontal and vertical deflection power was somewhat less than that for a conventional 25kv electron gun due to a divergent lens action. Proper shaping of the hat shield was employed to minimize the raster distortion which can result from such divergent lens action.

The contrast range of the tube was improved by putting the phosphor on the face plate. Also reflected primaries were somewhat reduced by a suitable choice of thickness of the aluminum backing.

Circuit developments have been mainly in the direction of more reliable and simplified RCS type operation, to permit full exploitation of the high brightness capabilities of the tube. In particular the arrangement illustrated in Figure 5 has been somewhat simplified by using balanced modulators as axis selectors which are gated at a 3.6mc rate (instead of a 7.2mc rate). This eliminates the frequency converter step, since the 7.2mc component can be directly selected from the harmonic content of the 3.6mc axis selected pulses, and also eliminates the need for generating the 10.7mc and 7.2mc signals. However, this simplification is accompanied by the possible disadvantage of requiring balanced modulators for axis selectors which, if they become unbalanced, can affect the color of the gray scale.

With the 1961 version of the tube having red as the unswitched color, the lower efficiency of the red phosphor was compensated for by using less switching. This provided approximately twice the dwell time on red as compared to either green or blue. With this arrangement good colorimetry was obtained particularly when adding a small fourth harmonic CW signal to the display to reduce the beam current during the transition time between phosphors.

s is for Shadow-mask

The shadow-mask color tube operates by masking a major fraction of the electron beam to prevent it from exciting other than the desired color phosphor. The color phosphors are deposited in dot trios aligned with corresponding holes in the shadow-mask - so that the angle of approach of the beam to the mask determines the color excited. It is almost always described as a three-gun tube but one-gun forms have been mentioned in the literature.

The Earlier Shadow-mask Tubes. By 1956 the shadow-mask tube had received a large development effort. The early tubes, using a flat shadow-mask with the phosphors on a flat glass plate mounted inside of the tube, had already been displaced by the curved aperture mask with the phosphors on the face plate.[29,30] The RCA 21AXP22 had been released, which was a 70° deflection angle tube having a metal envelope and having about 29 mils spacing between adjacent color dot trios. The color phosphors employed had unequal efficiencies and decay times. The reader is referred to the literature for further information on these earlier

tubes.

The More Recent Shadow-mask Tubes. The 21CYP22 described in 1958 contained a series of improvements although the overall size and deflection angle were about the same.[33] A glass envelope was employed (instead of a metal one) thereby simplifying the mounting problem even though the tube was heavier. The light output at the center of the tube was increased by use of a graded-hole shadow-mask in which the aperture size gradually increases toward the center compared to the edge regions. The contrast ratio of the tube was improved, particularly by using tapered apertures in the shadow-mask to reduce stray electron scattering, which could excite phosphor dots other than those intended. This tapered arrangement is illustrated in Figure 14.

The gun structure of the 21CYP22 was improved compared to previous designs particularly to reduce intercoupling between the convergence control effects - thereby easing the convergence procedure. The improved gun structure included an internal magnetic shield, as illustrated in Figure 15.

The more recently released 21FBP22 contained additional advances, particularly in the form of a new sulphide group of color phosphors. These new phosphors have more nearly equal decay times (the red and green being shorter than before), more nearly equal efficiencies and a more efficient red. This has resulted in less color smear on motion and more light output - as well as making equal drive operation more practical. Published data indicates that the light output for a given total beam current is about 70% greater (at the center of the screen) for the 21FBP22 than for the older 21AXP22.

The green phosphor of the new sulphide group produces a somewhat desaturated green compared to the previous green phosphor - resulting in some reduction in available color gamut. Further, the color of white recommended by RCA for alignment purposes is even bluer than before. This combined with the desaturated green results in a greater percentage of luminance from green and a smaller percentage of luminance from red and blue than with the NTSC standard primaries.

With the 21FBP 22 another improvement in the gun structure has been made, specifically by a brazed design to give better rigidity and provide a more stable cutoff characteristic for each electron gun.

The more recent versions of the shadow-mask tube like the 21FJP22 and 21FKP22 differ from the 21FBP22 by having an integral protective window. This eliminates the need for a separate protective glass in the receiver and thereby improves the contrast. In the 21FJP22 this protective window is filterglass giving a combined transmission of 43%, which further improves the contrast for a given ambient lighting condition.

Recently samples of a 23 inch rectangular 90° deflection angle shadow-mask tube were demonstrated. While it has been admitted that there are problems to be solved before such a tube is ready for mass production, the proponents of this tube feel that there are no unsurmountable problems.

During the period being reviewed, simplifications and improvements in convergence circuits have also been released. The convergence circuits now employed do not require the use of vacuum tubes, but are directly driven from the scanning circuits. In addition, rectifier elements have been included which insert a DC component to make alignment easier specifically by more clearly separating the convergence adjustments at the center and edges of the tube.

Another development in related components and circuits for the shadow-mask tube appeared in a recent production color receiver: namely, the use of permanent magnet centering. This eliminates the need for electrical centering controls by using a permanent magnet field which is effective at the deflection center.

z is for Zebra

The Zebra tube is an index colour television display, using photo-electric indexing, which was developed in England. It appears from the published information to be quite similar to the photo-index Apple tube. It employs an index structure having a periodicity other than the phosphor triplet periodicity combined with a non-ambiguous starting section. Further, it employs the UV radiation of a P16 phosphor (viewed by a photo multiplier tube) to provide the index signal. The reader is referred to page 523 of reference 17 for the picking of the name Zebra.

Other Circuit Developments

There are some circuit developments which appeared in more recent production color television receivers but which had been described prior to the period being

reviewed. These are noted here for general interest and they include: beam switching tubes for color demodulators, synchronous color killing, synchronous automatic chroma control, and the three triode matrix circuit (to obtain three color difference signals from a pair of color difference type signals).

Also during the period being reviewed there have been some technical papers on circuit developments which will be noted. One paper described a complete color decoder design using synchronous color killing and synchronous automatic chroma control practiced in a way which readily provides two mode color sync operation (for greater pull-in, with greater phase accuracy and less noise when pulled in).[36] Another paper described the use of magnetic demodulators for color television.[37] A third paper described a circuit arrangement in which the output of the R-Y demodulator is used to control the sub-carrier oscillator phase (APC) and the output of the B-Y demodulator is used to establish the chroma signal level (APC).[38] Yet another paper discussed various means for simplifying the viewer brightness and contrast controls on a color TV receiver by employing automatic circuitry, such as arrangements to prevent spot blooming and power supply overload.[39]

European Color System Studies

During these past few years various European color TV engineering groups have been separately evaluating the color TV situation and have been considering both other color display systems and other color transmission systems. The proposed Banana tube color display system has already been discussed in this paper. Two of the system arrangements receiving consideration will be noted here.

Modified NTSC

Simple modifications of the NTSC signal to conform to different vertical and horizontal scanning frequencies, and different picture-sound spacing (used in the various European systems) have been considered in the past. However, two additional modifications are currently receiving European consideration which should be noted.

The first of these modifications involves the use of different forms of gamma correction at the transmitter than the type in current use in U.S.A. The forms under consideration appear to be those which received preliminary study by the Gamma Subcommittee of Panel 13 of NTSC as discussed on page 213 through 216 of reference 2 and described in more detail in sections 9 through 15 of chapter 11 of reference 3.

The second type of modification being considered involves the use of the NTSC signal in a positive modulation system with AM sound. Such a system is claimed to be advantageous with regard to over-modulation problems (such as might exist with high luminance saturated colors) and with regard to the color-sound beat note problem by permitting the beat to always be interleaved and thereby less visible.

SECAM

This system (proposed by Henri De France) transmits a luminance signal (for the main signal) and has the two chrominance components (R-Y and B-Y) transmitted sequentially, rather than simultaneously (as NTSC).[40,41] The time sharing rate between the two chrominance components is every other scanning line - requiring the use of a storage element in the receiver (with a delay time of one scanning line and a bandwidth corresponding to that of the chrominance signal) to obtain simultaneous signals. The fundamental premise for such operation is that full vertical resolution is not required for the chrominance information.

This sequential transmission eliminates the variable phase aspect of the NTSC sub-carrier and the present SECAM proposal involves the use of a low amplitude FM'd sub-carrier (which therefore does not disappear on white). A claimed advantage for the system is that transmission vagaries, such as multipath and differential phase delay effects, produce less degradation than with NTSC.

The block diagram of a possible decoder circuit is shown in Figure 16. An ultrasonic type of delay line has been employed in experimental receivers for the storage element. To know in which direction the switch should be at any instant, some type of odd line synchronization is used. The present proposal involves the transmission of a color burst on every other line in order to "flag" a particular color difference signal.

Probably the main advantage for this system is an easing of the differential phase requirements upon intercity relay networks. Against this is the disadvantage of a reasonably high quality storage element in all receivers. The wisdom of such a compromise has been questioned in certain quarters.

References

General Color TV Theory

[1] J. W. Wentworth, "Color Television Engineering," McGraw-Hill Book Co., Inc., New York, N. Y.; 1955.

[2] NTSC, "Color Television Standards," McGraw-Hill Book Co., Inc., New York, N. Y.; 1955.

[3] The Hazeltine Laboratories Staff, "Principles of Color Television," John Wiley and Sons, Inc., New York, N. Y.; 1956.

[4] D. G. Fink, "Television Engineering Handbook," McGraw-Hill Book Co., Inc., New York, N. Y.; 1957.

[5] B. D. Loughlin, "Processing of the NTSC Color Signal for One-Gun Sequential Color Displays," Proc. IRE, Vol. 42, p 299; January 1954.

[6] R. G. Clapp, E. G. Clark, G. Howitt, H.E. Beste, E. E. Sanford, M. O. Pyle, R. J. Farber, "Color Television Receiver Design-A Review of Current Practice," Proc. IRE, p 297; March 1956.

Beam Indexing Tubes (Apple and Zebra)

[7] R. G. Clapp, E. M. Creamer, S. W. Moulton, M. E. Partin, and J. S. Bryan, "A New Beam-Indexing Color Television Display System," Proc. IRE, p 1108; September 1956.

[8] G. F. Barnett, F. J. Bingley, S. L. Parsons, G. W. Pratt, and M. Sadowsky, "A Beam-Indexing Color Picture Tube-The Apple Tube," Proc. IRE, p 1115; September 1956.

[9] R. A. Bloomsburgh, W. P. Boothroyd, G. A. Fedde, and R. C. Moore, "Current Status of Apple Receiver Circuits and Components," Proc. IRE, p 1120; September 1956.

[10] R. C. Moore, A. Hopengarten and P. G. Wolfe, "Techniques of Color Purity Adjustment in Receivers Employing the 'Apple' Cathode-Ray Tube," IRE Trans. Broadcast and Television Receivers, p 23; June 1957.

[11] J. B. Chatten and R. A. Gardner, "Accuracy of Color Reproduction in the 'Apple' System," IRE National Convention Record, Part 3, pp 230-237; 1957.

[12] D. Payne, H. Colgate, S. Moulton, C. Comeau and D. Kelley, "Recent Improvements in the Apple Beam-Indexing Color Tube," IRE National Convention Record, Part 3, pp 238-242; 1957.

[13] R. A. Bloomsburgh, A. Hopengarten, R. C. Moore, H. H. Wilson, Jr., "An Advanced Color Television Receiver Using A Beaming Indexing Picture Tube," IRE National Convention Record, Part 3, pp 243-249; 1957.

[14] D. E. Sunstein, "Index Signal Generating Means," U. S. Patent 2,892,123; June 23, 1959.

[15] I. Macwhirter, "Beam Indexing Tubes," Wireless World, Vol. 67, No. 1, pp 2-7; January 1961.

[16] I. Macwhirter, "Beam Indexing Tubes," Wireless World, Vol. 67, No. 2, pp 92-98; February, 1961.

[17] R. Graham, J. W. H. Justice and J. K. Oxenham, "Progress Report on the Development of a Photo-Electric Beam-Index Colour-Television Tube and System," Proc. IEE, Vol. 108, Part B, pp 511-523; September 1961.

[18] Technical Notebook, "Zebra Colour Television Display Tube," Wireless World, p 95; February, 1962.

Banana Tube

[19] "Banana-Tube Colour-Television Display," Wireless World, Vol. 67, No. 7, pp 351-352; July, 1961.

[20] P. Schagen, "The Banana-Tube Display System," Proc. IEE, Vol. 108, Part B, pp 577-586; November, 1961.

[21] B. A. Eastwell and P. Schagen, "Development of the Banana Tube," Proc. IEE, Vol. 108, Part B, pp 587-595; November, 1961.

[22] H. Howden, "Mechanical and Manufacturing Aspects of the Banana-Tube Colour-Television Display System," Proc. IEE, Vol. 108, Part B, pp 596-603; November, 1961.

[23] K. G. Freeman, "Circuits for the Banana-Tube Colour-Television Display System," Proc. IEE, Vol. 108, Part B, pp 604-612; November, 1961.

[24] R. N. Jackson, "Colorimetry of the Banana-Tube Colour-Television Display System," Proc. IEE, Vol. 108, Part B, pp 613-623; November, 1961.

[25] K. G. Freeman, and B. R. Overton, "Appraisal of the Banana-Tube Colour-Television Display System," Proc. IEE, Part B, Vol. 108, pp 624-630; November, 1961.

[26] P. Schagen, "Banana-Tube Color-Television," Electronics, pp 44-46; January 26, 1962.

Chromatron

[27] R. Dressler, "The PDF Chromatron - A Single or Multi-Gun Tri-Color Cathode-Ray Tube," Proc. IRE, p 851; July, 1953.

[28] J. D. Gow and R. Dorr, "Compatible Color Picture Presentation with the Single Gun Tricolor Chromatron," Proc. IRE, Vol. 42, pp 308-314; January, 1954.

Shadow Mask

[29] N. F. Fyler, W. E. Rowe and C. W. Cain, "The CBS-Colortron: A Color Picture Tube of Advanced Design," Proc. IRE, pp 326-334; January, 1954.

[30] H. R. Seelen, H. C. Moodey, D. D. VanOrmer and A. M. Morrell, "Development of a 21-Inch Metal-Envelope Color Kinescope," RCA Review, pp 122-139; March, 1955.

[31] M. J. Obert, "Deflection and Convergence of the 21-Inch Color Kinescope," RCA Review, pp 140-169; March, 1955.

[32] R. B. Janes, L. B. Headrick and J. Evans, "Recent Improvements in the 21AXP22 Color Kinescope," RCA Review, pp 143-167; June, 1956.

[33] C. P. Smith, A. M. Morrell and R. C. Demmy, "Design and Development of the 21CYP22 21-Inch Glass Color Picture Tube," RCA Review, pp 334-348; September, 1958.

[34] N. R. Goldstein, "The Effect of Several Variables on Phosphor-Dot Size in Color Picture Tubes," RCA Review, pp 336-348; June, 1959.

[35] J. M. Forman and G. P. Kirkpatrick, "Quality-Control Determinations of the Screen Persistence of Color Picture Tubes," RCA Review, pp 293-307; June, 1959.

Circuit Developments

[36] D. Richman, "Television Receiver Color Decoder Design," IRE Trans. Broadcast and Television Receivers, Vol. BTR-5, pp 27-45; January, 1959.

[37] M. Cooperman, "Magnetic Demodulators for Color TV," Electronics, p 56; January 2, 1959.

[38] Z. Wiencek, "Automatic Controls for Color Television," Electronics, p 58; May 15, 1959.

[39] L. L. Burns, R. W. Ahrons, L. B. Johnston, "Simplification of Viewer Brightness and Contrast Controls on Color TV Receivers," IRE Trans. Broadcast and Television Receivers, Vol. BTR-5, pp 54-66; May, 1959.

Secam

[40] R. Chaste, P. Cassagne and M. Colas, "Sequential Receivers for French Color TV System," Electronics, pp 57-60; May 6, 1960.

[41] R. Chaste and P. Cassagne, "Henri De France Colour Television System," Proc. IEE, Vol. 107, Part B, pp 499-511; November, 1960.

[42] "Colour Television N.T.S.C. v. SECAM," Wireless World, pp 515-516; October, 1961.

$$S_{NTSC} = Y + \alpha(R-Y)\cos\omega_c t + \beta(B-Y)\sin\omega_c t$$

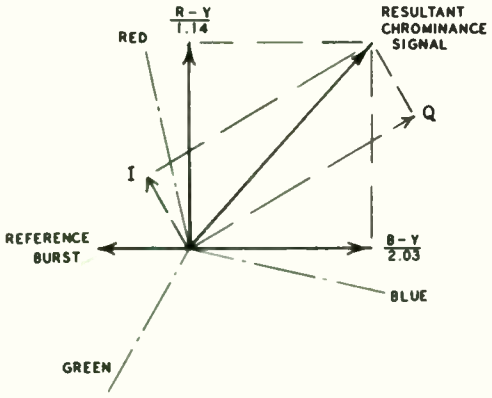
WHERE: $Y = 0.30R + 0.59G + 0.11B$

$$\alpha = \frac{1}{1.14} \quad ; \quad \beta = \frac{1}{2.03}$$

$$\omega_c = 2\pi f_c \quad ; \quad f_c = 3.58\text{MC}$$

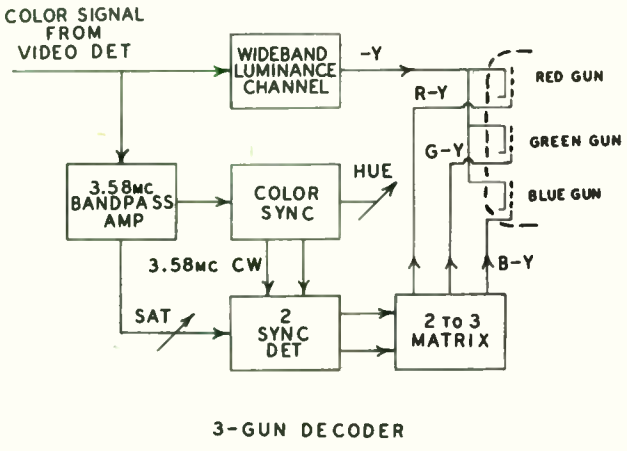
NTSC COLOR SIGNAL

Fig. 1.



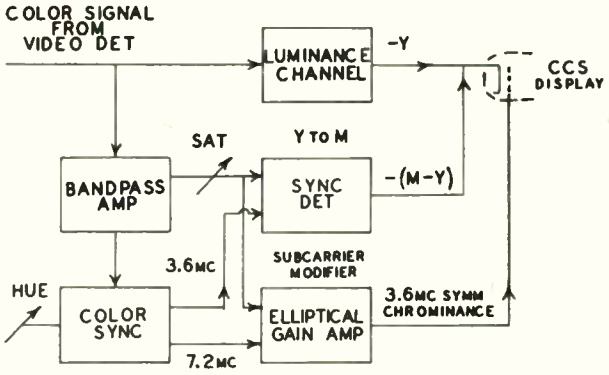
NTSC CHROMINANCE SIGNAL

Fig. 2.



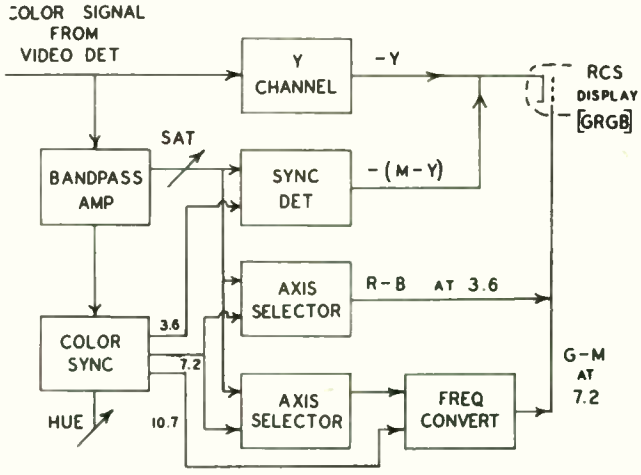
3-GUN DECODER

Fig. 3.



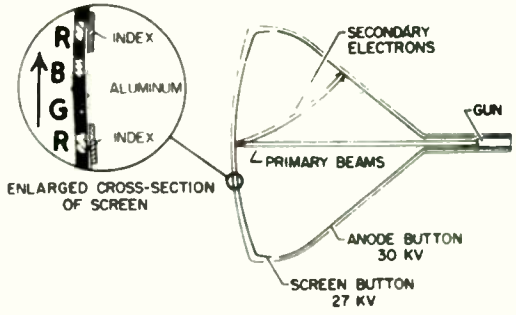
1-GUN CCS DECODER

Fig. 4.



1 - GUN RCS DECODER

Fig. 5.



SECONDARY-EMISSION INDEX APPLE TUBE

Fig. 6.

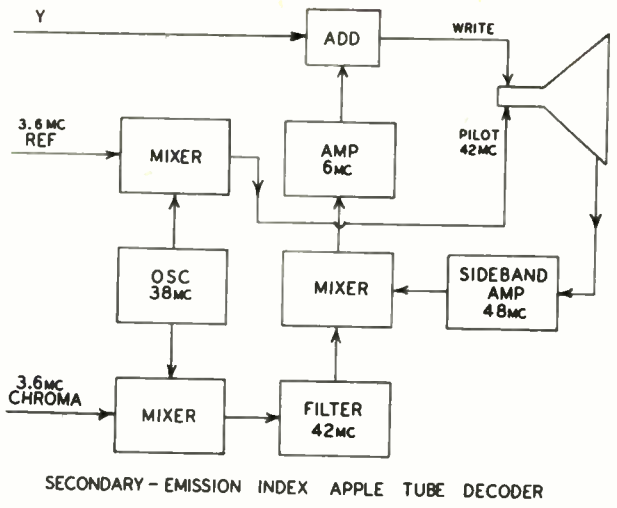


Fig. 7.

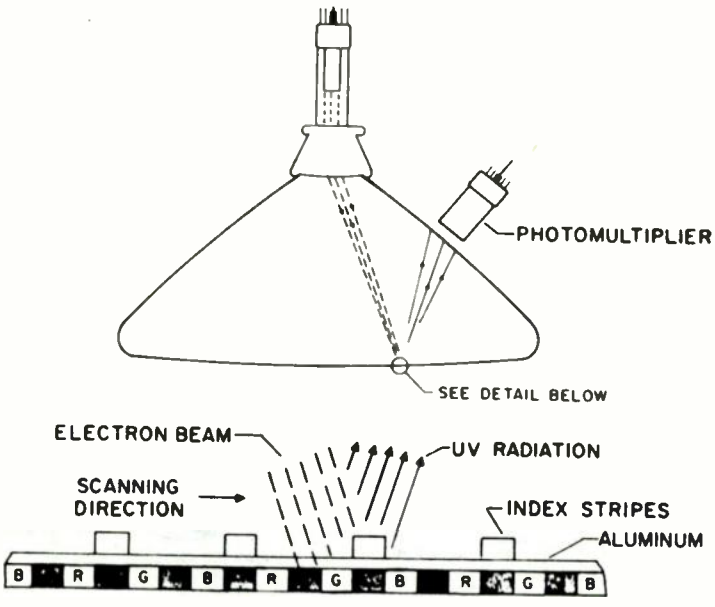
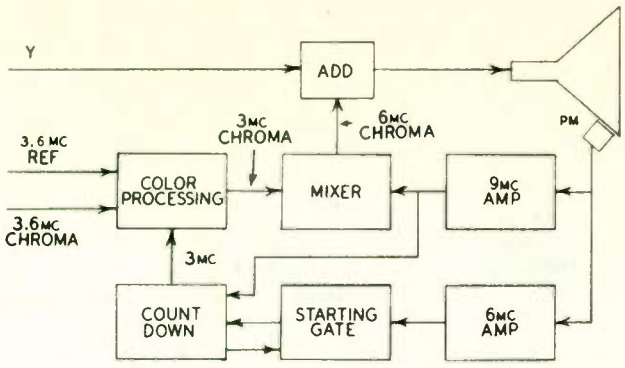


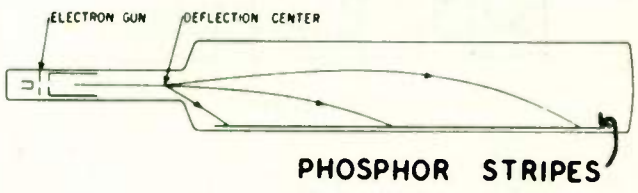
PHOTO INDEX APPLE TUBE

Fig. 8.



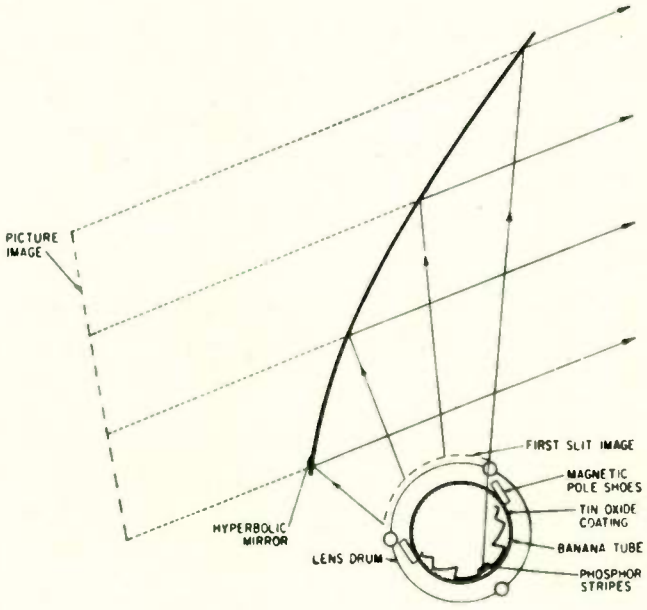
A DECODER FOR PHOTO INDEX APPLE TUBE

Fig. 9.



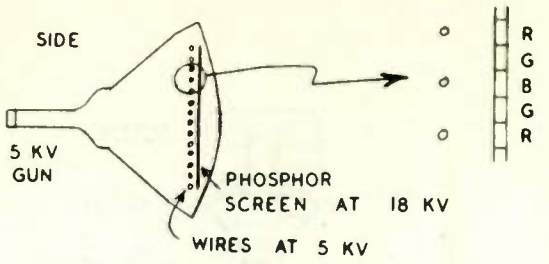
BANANA TUBE

Fig. 10.



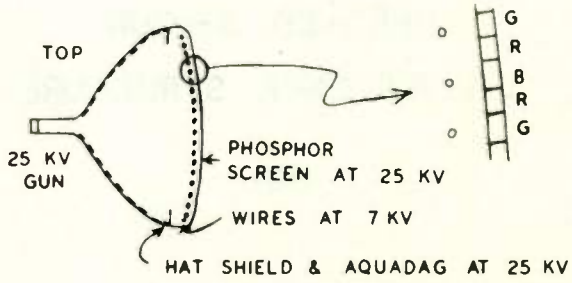
BANANA TUBE DISPLAY SYSTEM

Fig. 11.



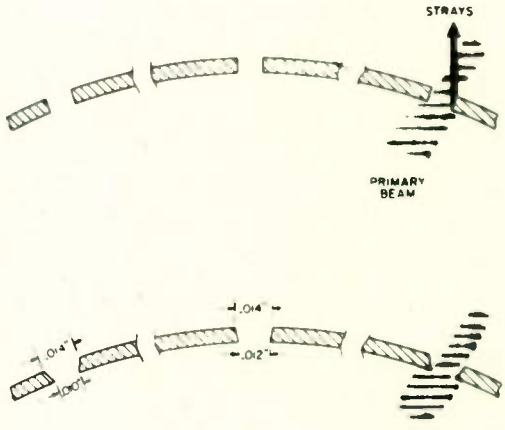
OLDER TYPE 1-GUN CHROMATRON

Fig. 12.



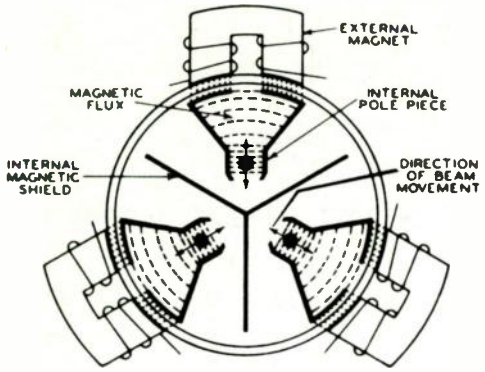
1961 1-GUN CHROMATRON

Fig. 13.



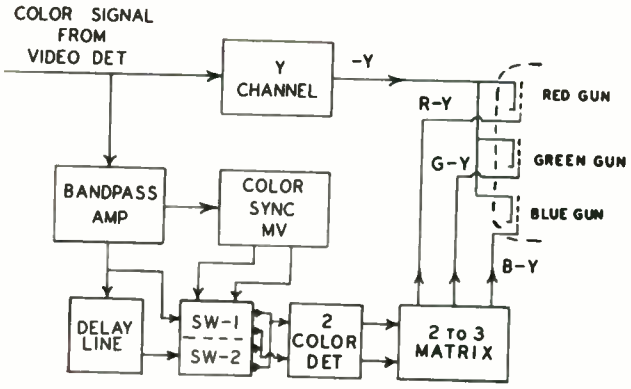
EFFECT OF TAPERED APERTURES IN SHADOW MASK

Fig. 14.



IMPROVED 3 – GUN CONVERGENCE STRUCTURE

Fig. 15.



SECAM DECODER

Fig. 16.

TELEVISION RELAYING VIA SATELLITES

H. E. Weppler
American Telephone and Telegraph Company
New York, N.Y.

The relaying of television programs by means of communications satellites gives promise of making world-wide television transmission feasible. This paper provides a qualitative review of a number of the considerations involved in such relaying. It is intended to provide general background information for engineers primarily versed in broadcast and television receiver techniques.

Why?

National Networks

The early history of radio broadcasting and, subsequently, the development of television broadcasting followed a pattern of spotty, limited growth as long as each station had to produce and finance all its own program material. The technical novelty was very real, but there were limits on the scope and quality of programs which resulted in only lukewarm public acceptance.

As the concept of network operation developed and means were developed to interconnect large numbers of stations over transcontinental distances, the consequent pooling of talent and programming funds provided the stimulus which triggered off the boom in public acceptance. In the case of television, this occurred in the late 1940's and early 1950's, using transmission facilities provided by the telephone company, just as in the case of radio networks.

Television programs have been originated from and delivered to virtually every part of the United States. In addition to the regular facilities used to tie together the many network stations, temporary combinations of facilities have been put together for a great many special occasions. The public has come to accept what often amounts to a small miracle of engineering and construction as commonplace; the opportunity to see almost any major event in the United States, as it occurs, is assumed to be practically an inalienable right.

At various times since the inauguration of television network operation there have been predictions that other means for programming would diminish the need for direct transmission facilities between the stations. Some even anticipated that the need for "live" networks would dwindle to just news and special events. You may be sure that such prospects have been of great interest to telephone people as they contemplated putting more and more millions of dollars into intercity facilities.

Kinescope recording, "live" films and, finally, video tape have each appeared. Each has served a useful purpose and has seemingly found its proper place in the scheme of things, but the demand for live networks has not diminished. In fact, the need has grown to the point where nearly 90,000 channel miles of television facilities are now being used.

The Dream

As national networking has developed to a fine degree, the dream of live intercontinental television has cropped up repeatedly. Coronations, Olympic games, cultural programs and even the Folies Bergere have been mentioned as typical attractions.

A vivid testimonial to the interest in intercontinental programming has been given in the great effort and expense which the networks have expended, in the absence of facilities for live relaying, to bring pictures with minimum delay on special occasions. There have been films and kinescope recordings flown by commercial planes and military jets. Special studios have even been set up at airports to speed up the handling on the ground.

Then too, there has been some work on a film process, sending pictures slowly, frame-by-frame, over the narrow band facilities of the submarine cables. A very laborious process to provide footage which will whisk through in a few seconds in the ultimate showing.

These efforts, and the appreciation of the public for them, demonstrate the inherent demand for intercontinental relaying. True, there are arguments that time differences, language differences and varying ethnic preferences limit the application. However, many of these problems also exist within continents, and certainly within Europe, yet the Eurovision network has come into being and has developed techniques to cope with the language problem.

Network television is not the only potential use for intercontinental television facilities. It is expected that today's intensive international trade, and the ever-quickening tempo of it, will produce many business uses for closed-circuit television. And its usefulness for defense is easy to visualize. Then there is also educational television. The composite of all these needs warrants serious attention to the problem of providing television facilities across the seas.

How?

Earlier Ideas

For many years overseas communication has been provided by means of high frequency radio, but h-f techniques just are not capable of providing the broad basebands needed for live television transmission. In recent years, underseas telephone cables have been introduced but their bandwidths are not yet of the magnitude required for television.

When the feasibility of over-the-horizon radio transmission at UHF and microwave frequencies became generally known there were various studies of possible trans-oceanic routes, island-hopping as it were. The problem of operating a number of very long links in tandem, particularly for such bandwidths, made this a difficult method to prove in and the early enthusiasm pretty well died out, at least as far as international television is concerned.

About the same time studies were made of the possibility of providing microwave repeaters in airplanes to bridge an ocean. Given enough dollars, a lot of airplanes and a fair sprinkling of luck such a system could probably be made to work. The proposal which probably came the closest to making sense economically would have put the microwave repeaters in commercial passenger-carrying aircraft. If a steady flow of aircraft could be maintained, with planes entering and leaving the "pipe-line" at the proper times, continuous microwave transmission would be feasible. The logistics of such an operation, to coordinate the transportation and communication services in all kinds of weather are staggering and it is hard to believe that any high degree of service reliability could be maintained. In any case, the scheme just hasn't looked sound enough to warrant going ahead.

Satellites Look Good

Perhaps one might philosophize that, having already considered methods operating below the surface of the ocean, along the surface, a few thousand feet above the surface and on up to the ionosphere, the next step should be to investigate methods using still higher altitudes.

The thinking has probably not followed this exact pattern but, in any event, the successes of the past few years in orbiting man-made satellites have served to focus attention on the possibility of relaying television programs by means of such satellites.

Satellites which are orbiting the earth at altitudes of at least several thousand miles will be mutually visible at points on the earth's surface which are thousands of miles apart. This by-passes the problem of finding strategically situated islands along the way and makes it technically feasible to provide broad band interconnections between practically any areas of the world.

Possibly a fortunate coincidence is that the concept of satellite communications has come at a time when the demands of international commerce and the public at large for intercontinental telephony are growing rapidly. It has become obvious that, where international telephony once could be accommodated with a few circuits here and there, the need now is for hundreds of circuits and in the not too distant future will need to be measured in the thousands. Services such as data transmission and voice-frequency telegraphy, which can be operated over telephone circuits, are adding to these requirements.

The way to provide such quantities of circuits economically is to multiplex them onto a broad band facility. Domestic experience has been that a broad band which is convenient and economical for large bundles of telephone circuits is of about the same bandwidth as that required for television transmission. Consequently, many of the studies and plans for satellite communication envisage satellites which may be used interchangeably or simultaneously for telephone and television purposes.

Cost

The major elements of a satellite communication system are far from cheap. The satellites, including the expense of launching, represent a million or more dollars apiece and the cost of ground stations is also measured in millions. However, with the opportunity to provide many hundreds of telephone circuits plus television with the same facilities the costs are spread out and the whole undertaking becomes a sound one over the long term.

No one is going to get rich quick in providing a satellite communication system. Expenses will be heavy right from the start but revenues will develop gradually, so there will be a number of lean years at the beginning. The best bet for economical operation and a sound over-all program is to lump the television requirements with the needs for telephone, data and related services and use satellites to extend the microwave techniques, now used over land, to long overseas paths.

The Outlook

The year 1962 is to be a big one in satellite communications, for both television and telephony. We are not talking today of mere dreams or paper plans. "Birds" with broad band repeaters inside are being built; some are in the final stages of preparation. Ground stations, too, are being built in the United States and by foreign partners. Within a few months we should see the first of a number of demonstrations of television, as well as telephony, via satellites. Before going into

further details of such an experiment it will be desirable to understand some of the technical factors involved.

Some Problems

Complexity

Although satellites are attractive for television relaying, it must be admitted that it will not be a simple job, and particularly since the economics are important. Not much connected with television has ever been easy. With the early television receivers it was a real task to get production quantities up and costs down to the point where sales could be widespread. Also, in the early days there was much conjecture as to whether such a complex receiver could be handled properly by the general public and whether, with so many components, it could be kept in working condition for any appreciable length of time.

The problem of relaying television signals, having a bandwidth of four megacycles or so, from place to place introduced many problems not encountered in the audio field, where five kilocycles of bandwidth takes care of most needs. When 100 microwave repeaters must be connected in tandem to cross the country the requirement for precision in each unit becomes very serious. In fact the very large number of vacuum tubes and other components which are involved, the failure of practically any one of which can interrupt the circuit, originally caused many doubts as to the practicability of such an undertaking.

In the end the problems have been solved, the doubts overcome and television has achieved the widespread acceptance once foreseen only by the dreamers. So it should be with satellites.

Transmission

The most difficult and expensive part of the satellite job is putting the payload into orbit, so the weight of the satellite and its contents must be very carefully considered. The source of primary power aboard the satellite will be a major factor in the over-all weight and, thus, the power requirements must be considered very carefully. Also, to be economical the satellite repeater must remain in orbit and continue to operate for a period of years.

Considering the need for light weight, low power requirement and long life one quickly reaches the conclusion that the satellite transmitter must be held to a low power; a few watts in current practice. But the transmission paths between ground station and satellite are thousands of miles long, as compared with thirty miles for a terrestrial microwave hop. So the basic transmission problem becomes one of power conservation.

In microwave practice it is common to pick up gain in the transmission path, without adding active elements, by means of directional antennas. In the case of the satellite proper this technique is sorely limited. If simultaneous transmission in many directions is required, then an omnidirectional antenna is needed. If transmission is needed in only one direction, the use of a directional antenna will give gain but requires some means to accurately point it and keep it on the desired target at all times; this can be a very difficult requirement.

To arrive at a working system we must provide all the gain possible at the ground stations. The transmitting power must be very high by microwave standards, typically several kilowatts. Then a highly directional ground antenna provides a substantial boost.

The signal received from the small transmitter on the satellite is so weak that the most sensitive receivers known in the art are required. The best uses a maser amplifier, the inherent noise of which is very close to zero, so there is the least chance of masking the weak signal with internal noise. In some cases parametric amplifiers are used but the noise performance is not quite as good as the maser.

The low noise capability of the maser will be wasted if the antenna system picks up extraneous noise along with the desired signal. Hence, highly directional receiving antennas serve two purposes; to provide receiving gain and to minimize noise pick-up. Even highly directional antennas usually have minor lobes extending to the side or rear and these must be carefully controlled to minimize noise pick-up.

The horn reflector antenna appears particularly advantageous with respect to noise pick-up. One with an aperture 3600 square feet has a gain of 50 to 60 db and a beam width of about two-tenths of a degree. In practice, the same antenna may be used for both transmitting and receiving, with a diplexer to separate them. Parabolic antennas are also being used in satellite work in sizes up to 85 or 100 feet. The use of high-gain antennas is not without headaches; they become massive structures accompanied by all the problems that go with heavy construction. Because of their extreme directivity, very accurate control of pointing is required, while the inherent mass plus its wind loading make this difficult.

In considering power conservation, the method of modulation must also be especially selected to give the best performance in overcoming noise. At least for the early years frequency modulation is the first choice. To extend the breaking point, which normally would deteriorate in an FM system as the noise advantage is increased, a complex form of FM detection with loop feedback (FMFB) is being used. The feedback path controls the beating oscillator in accordance with the received signal to reduce the effective deviation and permit a comparatively narrow i-f pass band to be used.

Life

Long life for the satellite is an economic necessity. Hence, everything that goes into it must be prepared to give long service in a hostile environment. The equipment must also be able to withstand the stresses that go with launching. Shocks, spins and vibration during this phase are very severe. Of course semi-conductors are used almost exclusively; traveling wave tubes are probably the only vacuum tubes considered.

Much has yet to be learned about the effects of particle radiation in space on the life of semi-conductors and the early satellite experiments are being arranged to provide extensive data. In an operating satellite it is necessary to have telemetering channels to transmit information to the ground station as to the condition of critical units or components. Additional telemetering channels will be provided in the early units to provide even more extensive experimental data. A command channel from ground station to satellite serves to turn the electronic gear on and off to minimize radio interference and, at least in the early models, to conserve battery power.

Launching

Designing and building the rocket and then launching and guiding it to place the satellite in the desired orbit present a separate series of challenging and complicated tasks. Reliability, safety, accuracy and cost are all important. The communication engineer has no illusions or desires to "do-it-yourself". For these specialized skills he turns to the aerospace industry and so the communications industry becomes a customer of the rocket people; as it were, giving them a package to be delivered to the desired address.

Many Choices

The Parameters

In the design of such a complex communications system there are many parameters to be determined, many choices and trade-offs to be made. A number of the areas have been mentioned above. Practical experience through working experiments is needed before the final parameters for a full-scale world-wide system can be set down.

Several possible configurations have been described by different organizations. Each such plan has its advantages and disadvantages and its own particular definition of objectives. There has been a tendency for observers to assume that the respective organizations are heavily committed to their particular proposals in exact detail. In reality, the systems described should be looked upon as examples, in which substantial changes may be made as studies proceed and, particularly, as experience is obtained.

The one parameter which has generated the most interest in the proposals to date is the choice of altitude for the satellite.

Satellite Altitude

The altitude at which a satellite orbits the earth determines, of course, the extent of the area on the earth's surface from which the satellite may be seen at any given time. The higher the satellite, the greater the area (with an area equal to half the earth's surface as the theoretical maximum) and the more points which will be able to view the satellite at the same time for communication relaying. The thrust needed to put the satellite into orbit naturally goes up with altitude and so this has a heavy bearing on the cost and even on the availability of suitable launching rockets.

Medium Altitude. An altitude of about 7,000 statute miles offers a number of attractions. It is high enough to provide a substantial coverage area, in fact, it amounts to 63% of the theoretical maximum. It falls in a zone where the radiation intensity should be substantially less than at lower altitudes and so stands to limit the radiation effects on the semi-conductor components. And putting satellites of the mass expected for communications into such an orbit will not require rockets of extreme thrust capabilities. In fact, it will probably be feasible to launch several satellites from one rocket and thus improve the overall costs.

Such satellites will orbit the earth about every six hours, so from any given ground location they will travel across the sky, rising and setting like the other celestial bodies. The ground station antennas must be able to follow the satellite across the sky, and it will be necessary to have a number of satellites in orbit, at different positions, to provide a high degree of availability. Up to fifty satellites have been proposed, the exact number depending on the particular job to be performed.

The techniques needed for satellites at such altitudes are well under control now and it should be an orderly, predictable procedure to achieve a working system.

Stationary Type. Another configuration receiving a lot of attention is the "stationary" satellite. At a specific and much higher altitude, near 22,000 miles, a satellite moving eastward in an equatorial plane will remain stationary, or nearly stationary, above the earth's surface. Then a given satellite may be used continuously, rather than intermittently, by a pair or group of ground stations and the ground station antennas need be steerable only to the extent necessary to correct for drift. Of course one satellite isn't enough for a

working system; three are required to cover the globe but, beyond that, each satellite should have a spare nearby to pick up the load in case of failure. Looking farther into the future, as the total communication requirements build up it would be necessary to divide the load among a number of separate stationary satellites in order to keep the radio spectrum requirements within practical bounds and to ease other technical problems.

Major problems inhibit any early decision to base working system designs on the stationary satellite. First, the problem of launching the satellite into the desired orbit and keeping it accurately on position is a difficult one. Rockets with sufficient thrust will be available in due time, but not as soon as rockets adequate for lower altitudes. Control techniques will undoubtedly be developed, but position control has not yet been achieved for anything like the lifetime desired for a communication satellite.

With a satellite at 22,000 miles the time required for the radio waves to travel up to the satellite and back down, even though traveling at the speed of light, amounts to several tenths of a second. This should not be a serious problem for television and other one-way services, including at least some types of data, but such delays may be serious for telephony. The problem is being studied intensively and conclusions as to the effects on normal telephone usage, rather than just isolated tests, will need some more time. This matter is of importance to the television application, at least in the early stages, because television and telephony will probably share the same satellites.

Status. Work is going on toward both types. Some experts have even proposed concentrating on the stationary type to the exclusion of others. However, it is interesting to note that none of the experiments now scheduled is to achieve a truly stationary orbit and apparently no dates have been set for this achievement. This gives strong confirmation to the wisdom of the plans for lower altitude systems to initiate satellite communications. In the long pull, we may well see a combination of the two.

Direct Broadcasting from Satellites

Another Way

The foregoing has all referred to the use of satellites to relay television programs to or between conventional television networks. The final link in that chain is the broadcasting of the program to the public's receivers by conventional television broadcasting stations.

Direct broadcasting from satellites to the public has also been mentioned in some circles. When you stop to think it through, this turns out to be a mighty difficult proposition. To use the same techniques as those being planned for communications satellites, it would be necessary for each receiver to incorporate all the features of the ground stations described earlier. Giant dish or horn antennas and the rest of that gear are patently unreasonable in both complexity and cost for home use. However, all the basic problems in transmitting broad band signals over very long paths economically are still applicable.

High Powered Satellite

To simplify the individual receiver installations will require (1) the use of high power in the satellite, to provide a stronger signal at the earth's surface, and (2) the use of stationary satellites so that the individual receiving antennas do not have to track the satellite across the sky.

Higher power in the satellite does not mean just a few more watts output from the satellite transmitter, but several orders of magnitude more power. Figure 1 shows some possible characteristics for such a project. This assumes the use of ordinary home receivers, with vestigial sideband transmission in the UHF band.

If the satellite used an antenna of moderate directivity, say one with 20 db gain, it would permit coverage of at least one-third of the earth's surface. Assuming the home receiver is a conventional UHF receiver and is equipped with only a simple dipole antenna, in order to provide the field strength specified by the F.C.C. for Grade B coverage the satellite output power would have to be about 1,000,000 watts. The difference between a megawatt and the two or three watts proposed for satellite relaying represents the difference between a conventional home receiver with a small antenna and a ground station with a maser receiver and a 3600 square foot antenna aperature. This, in turn, represents the difference between a $200 receiver and a receiver costing about 10,000 times as much.

Now various trades can be made. At the expense of bigger antennas for each receiver, say a six foot dish, the power could be reduced to 20 kilowatts. Parametric receiving amplifiers might be assumed and perhaps a picture quality below Grade B could be tolerated. But these are questionable concessions and still do not bring the satellite power requirement down to a level which is currently feasible.

It would probably be naive to say that such power on the satellite can not be achieved, but it certainly will not be achieved soon. For one thing it will be necessary to have a power source other than solar cells. Nuclear power may be the answer. Then, it will require a mighty rocket to put the weight of such a transmitter and its power supply into the 22,000 mile orbit. Ultimately one nuclear supply might care for both the propulsion and electrical power needs, but this will not be light or soon.

As to the matter of cost, a high power, heavy satellite will be expensive to begin with and the possibility of insuring long life when operating at high power levels is not encouraging. Certainly we do not today operate high power stations without people, without adjustment, and without replacement of components for years on end.

So it seems safe to say that, for some time to come, the usefulness of satellites to the television industry will be in the relaying of programs. If there is to be direct broadcasting from satellites it will come much later.

Other Factors

Where direct broadcasting is intended to serve several countries at once, there will be problems as to transmission standards (scanning rates, etc.), languages, local tastes, financing and others. Some of these problems also apply to satellite relaying between countries but in the latter case they are much easier to cope with because conversions or adaptations can easily be made in the single relaying path which delivers the program to a given country. However, these are probably better subjects for discussion by programming or administrative people than by radio engineers.

One potential problem of concern to radio engineers is that of frequency allocations. Of course this will require international agreement. The use of satellites by one nation for direct broadcasting to people in other areas would give the originating nation such direct access to other peoples that some nations would probably object to setting aside frequencies for this purpose.

Programs relayed between networks would give each country the opportunity to determine the extent of its own participation. Idealistically the latter may not necessarily be superior, but it probably affords a better practical basis for international agreement on frequencies.

An Example

Project Telstar

Of the several communication satellite experiments scheduled for 1962, Telstar is the one with which I am most familiar and so I shall describe that briefly as an example of the work going on. Project Telstar is being sponsored by the Bell System. The rocket and the launching will be provided by the National Aeronautics and Space Administration at Bell System expense while the satellite itself and the broad band ground station in the United States have been developed and are being built by the telephone company.

Over-all Plan

Project Telstar will provide the means to measure the transmission performance of a broad band satellite under actual orbital conditions. The satellite will contain one broad band repeater and will be able to relay one television signal or the equivalent of about 600 telephone channels in either direction, to or from the United States. (A fully operational satellite would have two such repeaters for two-way transmission.) Of course, the 600 one-way telephone channel capability of the experimental satellite will not be useful operationally, but it will permit the many technical tests necessary to test the repeater's operating characteristics. For tests of two-way telephone transmission the broad band of the satellite will, in effect, be divided into two narrower bands. Using one such sub-band in each direction will permit operational tests of 12 to 60 simultaneous two-way telephone channels.

Equally important will be the operational testing of the related ground-based equipment, including the means for satellite acquisition and tracking.

This opportunity will also be used to learn more about the space environment and its effect upon semi-conductor devices and other components. Information as to particle radiation flux, ambient temperature and component performance will be telemetered back to the earth stations.

The Thor-Delta rocket will be used for this launching because of its earlier availability and because of the background of successful experience in building and launching this type of launch vehicle. Adjusting the orbit within the capabilities of the rocket, and to obtain as much useful data as possible, the plan is to achieve an elliptical orbit at an inclination of about 45 degrees to the equator. The highest altitude of the ellipse is to be about 3500 statute miles and the low point should be about 575 miles. In this way, the satellite will pass in and out of the high intensity region of the Van Allen radiation belt, giving much more scientific data than if the orbit were circular at an intermediate altitude.

Such an orbit will have a period around the earth of about two and a half hours. This will provide mutual visibility between the United States and Europe for periods up to a little over a half-hour for individual passes and up to one-and-a-half hours total per day. The usable periods between two tests points in the eastern United States would, of course, be longer. The mutual visibility periods are not the same day after day because, in addition to the earth's motion under the satellite, the axis of the elliptical orbit will progress around the earth on a cycle of about 180 days. It is felt that the useful periods will be more than sufficient for the full testing program. Of course, a fully operational system would use a more nearly optimum orbit and sufficient satellites for a high degree of availability.

Radio Frequency Plan

The broad band communication channel through the satellite will be transmitted from ground to satellite at about 6390 mc. Obviously, a different frequency must be used from satellite to ground to prevent singing; this will be at 4170 mc. These frequencies are both in bands used for common carrier terrestrial microwave systems.

The satellite repeater will be a linear transponder, that is, it will amplify the incoming signal and shift its frequency without any change in the modulated form of the signal. Its bandwidth will be about 50 mc.

For two-way telephone tests, one set of relatively narrow band signals will be about 5 mc above the center frequency of the repeater and the other set 5 mc below the center.

Auxiliary signals are needed for tracking, telemetry and control. For the broadest measurements of the satellite's location, a 136 mc beacon aboard the satellite will radiate continuously. This will be observed by NASA's Minitrack network as well as the communications ground stations. For more refined tracking a 4080 mc beacon will also be radiated. Telemetry from the satellite will be superimposed on the 136 mc beacon. Command transmitters, on the ground, at about 120 megacycles will send orders to the satellite.

The Satellite

The satellite proper, as seen in Figure 2, will be a sphere, about three feet in diameter. It will weigh about 170 pounds, complete with all antennas and electronic gear. The greater part of the surface will be occupied by some 3600 solar cells to provide the primary power. Around the middle of the sphere are two belts of antenna apertures, constituting the 6 kmc (gc) and 4 kmc (gc) antennas. Shown on top is the multi-spiral VHF antenna which is set into position when the satellite reaches orbit and the nose cone drops away. A VHF antenna for temporary use before extension of the regular antenna is located on the underside.

The microwave antennas and multi-spiral VHF antenna have radiation patterns very nearly isotropic. Hence, close orientation of the satellite is not needed but a degree of stabilization will be provided as a result of gyroscopic action resulting from a spin imparted by the launching vehicle.

The microwave transmitter uses a traveling-wave-tube in the ouput stage and develops an r-f power of two watts. The tube is based on earlier designs which have demonstrated long-life characteristics and the ability to withstand the mechanical shocks of launching. The balance of the electronic gear uses semi-conductors; the total, including the telemetry, command and beacon functions, comes to 1300 diodes and 1000 transistors. The design objective for the experimental unit is a working life of two years.

Ground Station

The principal ground station in the United States will be near Andover, Maine. It has been necessary to go to this remote area because of the need to share the common carrier 4 and 6 kmc bands. Line-of-sight microwave links operating at 11 kmc will connect the site to the nationwide broad band facilities.

It will have a 3600 square foot horn-reflector antenna, a structure 177 feet long, 94 feet high and weighing 315 tons, shown in a cut-away view in Figure 3. This will be enclosed by a giant radome. The high degree of directivity of the antenna may be grasped by understanding that the antenna's beam will be only 13 miles wide at the satellite when at its maximum altitude.

In cabs attached to the rotatable antenna will be the two-kilowatt broad band transmitter, maser receiver, and many auxiliaries.

An over-all view of the station is shown in Figure 4. The antenna radome is seen in the foreground. The control building at left center houses the telemetry, command and tracking electronics, the video and audio terminals, general supervisory and operating facilities and an emergency power supply.

The site has been planned for the addition of more large antennas to ultimately provide a fully operational station capable of communicating with several satellites simultaneously.

Operation

Project Telstar is to be launched the second quarter of 1962 and communication experiments will, of course, start at once. Ground stations are under construction now in England and France. Transatlantic tests will start as soon as those stations are ready. In the meantime, much useful testing can be performed by means of loop transmission from Andover to the satellite and back to Andover. Tests will also be made from Andover to a receiving-only station at Holmdel, New Jersey.

Conclusion

The use of satellites for relaying television and telephony, including data and related forms of communication, seems assured. Not all the parameters have been buttoned down as yet, but the experiments scheduled for 1962 will go a long way toward providing data on which to make those determinations. And they will give the public its first look at live intercontinental television.

The description of Project Telstar has only

outlined the major elements of the first experiment. Other experiments will follow and will gradually incorporate greater sophistication. The implementation of a fully operational system should follow in orderly progression.

FREQUENCY	700 MC (CHAN. 52)
SATELLITE TYPE	SYNCHRONOUS (22,300 MI.)
SATELLITE ANTENNA GAIN	20 DB
COVERAGE AREA	1/3 OF EARTH'S SURFACE
TYPE OF SIGNAL	U.S.A. STANDARD
RECEIVER	CONVENTIONAL UHF
RECEIVED SIGNAL	F.C.C. GRADE B
RECEIVING ANTENNA	DIPOLE
SATELLITE TRANS. POWER	1 MEGAWATT

OR

RECEIVING ANTENNA	6 FOOT DISH
SATELLITE TRANS. POWER	20 KILOWATTS

Fig. 1.

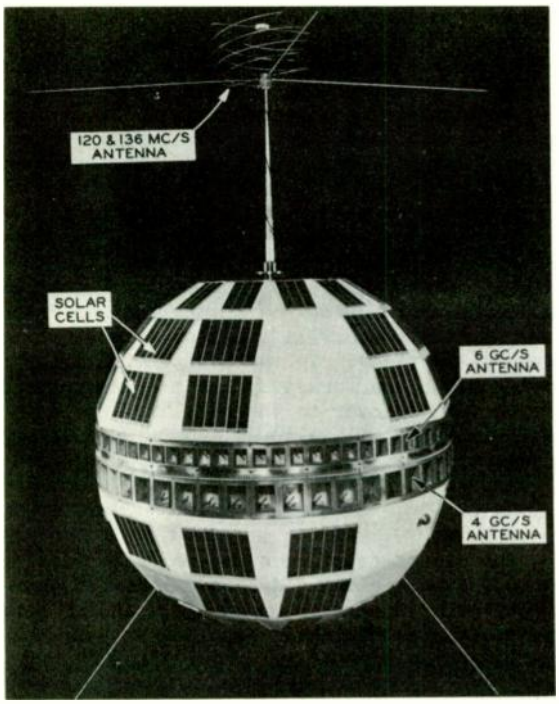


Fig. 2.

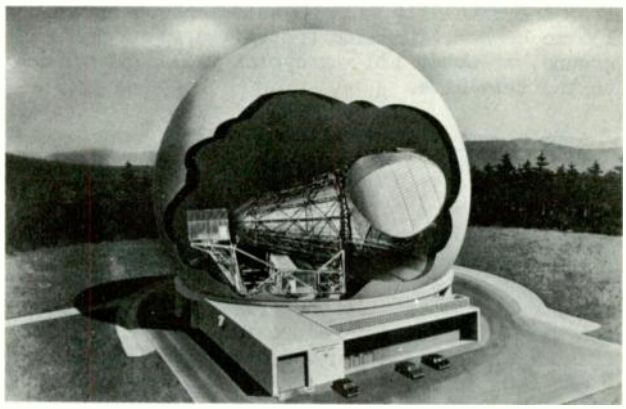


Fig. 3.

Fig. 4.

THE FUTURE OF TRANSISTORS IN TELEVISION RECEIVERS

Roger R. Webster, Texas
Instruments Inc., Box 5012,
Dallas, Tex.

Abstract

If transistors are to successfully compete with vacuum tubes in television receivers, they must do it on price and performance. The sales appeal of "portable" operation is insufficient for a mass invasion of the market.

There is little question that transistors can compete performance-wise; the problem is, will transistors compete at comparable cost? Transistors may receive some premium, particularly at the outset, since it appears likely that the extra reliability of transistors will reduce warranty costs.

This paper reviews these factors in relationship to performance. Particular emphasis is placed on horizontal deflection circuits and front end design. Some educated "guesses" as to the near future of transistors in T. V. receivers are made.