THE BELL SYSTEM
TECHNICAL JOURNAL

VOLUME XXXIV MAY 1955 NUMBER 3

A Revised Telephone Transmission Rating Plan

By P. W. BLYE, O. H. COOLIDGE, and H. R. HUNTLEY

(Manuscript received January 25, 1955)

Although the telephone appears to be a relatively simple device, it is in fact one of the more complex elements of the telephone business, primarily because its operation depends on transforming sound waves into electrical waves and vice-versa. This fact makes its performance difficult to evaluate and measure. It is not surprising, therefore, that the problem of evaluation and measurement, that is, of "rating", has been the subject of much thought and work over many years and that even today widely different methods are used in different parts of the world.

The Bell System has done its full share of work in this connection and as the art and state of knowledge have changed, several different rating methods have been used. In the past twenty-five years or so the methods have come almost full circle from "loudness" through "effective transmission" to "loudness" again; but with a significant difference. The earlier ratings (loudness originally, effective later) were always in terms of so many db poorer or better than a standard rating. While this standard was specifically defined, only a few people were sufficiently familiar with it to appreciate thoroughly whether something which was rated, say, 6 db better than the standard was good, bad or indifferent.

The new rating system avoids this problem by not having any fixed standard at all. It is based primarily on the ratio of the loudness of the sounds into the listener's ear to the loudness of the sounds out of the talker's mouth but with suitable modifications to take care of factors other than loudness which affect ease of conversation.

This system also has the advantage that "gain" and "loss" have acceptable connotations; that is, an over-all gain means that the weighted sound pressures out of the receiver are greater than the weighted sound pressures at the talker's lips and a loss means the reverse. Thus this system should not be rendered out of date by any foreseeable developments in subsets or other transmission instrumentalities.

While the quantities used in the new rating method strictly speaking are not decibels, it is expected that the adoption of this method will have no effect on the usage of the term "decibel" either inside or outside of the Bell System.

This article discusses the over-all problem, gives a brief historical review and describes in some detail the new rating system now being adopted.

INTRODUCTION

In the years prior to about 1930, performance ratings were assigned to telephone transmission circuits in terms of loss of speech volume or loudness. The adverse effects on transmission quality of such parameters as sidetone, distortion and noise were recognized, but until the 1930's no way was known of incorporating them, together with loudness loss, into a single figure of merit for rating transmission performance.

Anti sidetone station sets were first used in quantity in the Bell System in 1932, followed in 1937 by the 302-type combined sets. While these instrumentalities did not yield appreciably greater loudness than their predecessors, they exhibited marked improvements in such characteristics as sidetone, frequency response and nonlinear distortion, which have a marked effect on the ability of a listener to *understand* as well as to *hear* transmitted speech. Thus it became necessary to devise a rating system which would give adequate recognition to the contributions of these improvements to over-all transmission performance. There resulted the "effective loss" method of transmission rating.[1, 2]

The effective loss rating system assigned a figure of merit to a customer-to-customer transmission circuit, based fundamentally on the rate at which listeners requested repetition of what the talkers had said. This basis evaluated transmission circuits from the standpoint of transmitted intelligibility, as well as loudness. In practice, the circuit to be rated was compared with a "working reference system,"[2] which consisted of representative subscriber loops and station sets, and a variable, distortionless

[1] W. H. Martin, Rating the Transmission Performance of Telephone Circuits, B. S. T. J., **10,** p. 116, Jan., 1931.

[2] F. W. McKown and J. W. Emling, A System of Effective Transmission Data for Rating Telephone Circuits, B. S. T. J., **12,** p. 331, July, 1933.

trunk. In principle at least, the trunk of the reference system was adjusted until the repetition rate was the same on both circuits. The attenuation of the trunk in decibels was then a measure of the effective loss of the circuit under test.

An effective loss of zero on the working reference system was a theoretical point, arbitrarily assigned so that the system tied in numerically with its predecessor. If a listener could not hear *and understand* a given talker as well over the circuit being rated as over the reference system at its zero point, the circuit under test was said to have a *positive* effective loss. If the reverse was true, the circuit was said to have a *negative* effective loss. Effective loss was expressed in db and, from a practical point of view, may be considered an *equivalent* loss in loudness.

While the effective loss system has served an important purpose in giving due weight to important transmission characteristics other than loudness efficiency, it has a number of practical disadvantages. For example, the effective losses of circuits containing electro-acoustic transducers, such as telephone transmitters and receivers, cannot be measured by objective methods. Hence they must be determined statistically by time consuming subjective methods such as repetition tests. Relatively few people have had the experience, personally, of listening to speech over the reference system; hence there is little tie-in between effective transmission and practical experience.

Also, the telephone transmitters and receivers used in the reference system were of obsolete types, their frequency characteristics were far from flat and the circuits in which they were used introduced both sidetone and nonlinear distortion. Hence, modern station sets invariably exhibit performance superior to those used in the reference system and thus yield negative effective losses. The combination of positive and negative losses, together with the inability of the practical user to correlate the net result with any physical bench mark with which he was familiar, often prevented complete understanding and acceptance of the rating system, except by transmission specialists. Although transmission engineers have become expert in designing exchange area plant based upon effective loss techniques, the fact that effective losses cannot be measured directly prevents close correlation between design and maintenance. This is an important drawback because wherever practicable it is very desirable to be able to measure performance of all parts of the plant in the same terms that are used in laying it out.

The telephone art has now advanced to a point where a much simpler rating system appears practicable. It is the purpose of this paper to describe the more significant features of a modified system which, it is

believed, will eliminate most of the practical disadvantages of the effective loss plan.

PROPOSED RATING PLAN

As pointed out above, the effective loss method of rating transmission circuits was adopted in order to give proper recognition to factors such as sidetone, frequency response and nonlinear distortion, which, in combination with loudness, determine the over-all performance of a telephone circuit. Today, between 80 and 90 per cent of Bell System telephones provide reasonably effective antisidetone features and it seems evident that on the average little additional effective gain can be accomplished by further reducing sidetone. Also, in the case of the modern 302- and 500-type station sets, which constitute a constantly increasing portion of the station plant, impairments due to frequency response and nonlinear distortion are so small that they leave little need for improvement in future designs. It seems evident, therefore, that loudness has again become the main variable factor in plant design and is likely to be the most important consideration in future telephone set design. In view of this situation it seems desirable to return to a rating system based primarily upon loudness considerations.

While one of the chief reasons for returning to a loudness basis is the need for a rating system in which the parameters can be measured by objective methods, it is recognized that, strictly speaking, loudness itself is a subjective factor which can be measured only by the composite judgment of a number of human observers. As used here, however, loudness is determined by objective measurements which do not always duplicate true loudness determinations precisely, but which are sufficiently good approximations for practical purposes.

Although it is contemplated that *loudness* loss will be used in the great majority of day-to-day engineering problems, other factors contributing to effective loss will not be disregarded. In special cases, where these may assume importance, allowances may be made for them in the form of "penalties" to be added to *loudness* loss. The resulting rating, including all penalties, will be referred to as *subjective* loss to distinguish it from *loudness* loss and the present *effective* loss. It is contemplated that the use of subjective losses will be necessary only for some special service applications, for the design of apparatus and systems, and for cases in which noise is important. It is recognized that occasionally there may be special types of speech transmission systems in which factors other than loudness are controlling. In such cases it may be more practicable to

provide ratings by means of over-all subjective tests rather than by measurements of loudness corrected by the penalties referred to above.

To provide maximum usefulness in the engineering of the telephone transmission plant, a rating system must include means not only for appraising an over-all telephone connection but also for assigning figures of merit for the various components making up the connection. Thus it is highly desirable that it make possible the rating of transmitting circuits, connecting lines, and receiving circuits as separate entities.

Definitions covering the rating of component portions of a telephone system must, of course, be consistent with the definition for over-all system rating. It becomes necessary, therefore, to specify under what conditions the sum of the component ratings is equal to the over-all rating; other conditions require reflection corrections in order that the rating system be consistent.

The rating definitions involve both acoustic pressures and electric voltages. They are based on the use of speech, or an equivalent complex voice frequency test tone and artificial mouth, as a source of acoustic energy, and on meter indications of voltage or sound pressure suitably weighted to simulate the loudness perception of human hearing.

OVER-ALL LOUDNESS RATING

A loudness transmission rating of zero is assigned to an over-all telephone connection in which the output acoustic speech pressure delivered by the telephone receiver to the ear of the listener is equal to the input acoustic speech pressure at the lips of the talker. The measurement of talking pressure is made under free field conditions, that is, with the telephone transmitter removed from in front of the talker's lips or other sound source. During the measurement of output pressure, the transmitter of the talking telephone set is placed in the modal position with respect to the sound source.[3]

As shown in Fig. 1, if S_L is the output pressure at the listening end in microbars (dynes per square centimeter) or millibars (thousands of dynes per square centimeter), and S_T is the corresponding input pressure at the talking end of an over-all telephone connection, the transmission loss, or rating, of the connection, R_0, is

[3] The modal position is the same distance from, and angular position with respect to, the sound source as the transmitter would assume when used correctly by a talker whose facial measurements are the mode of a very large distribution representative of telephone subscribers. See W. C. Jones and A. H. Inglis, The Development of a Handset for Telephone Stations, B. S. T. J., **11**, p. 262, April, 1932.

$$R_0 = -20 \log_{10} \frac{S_L}{S_T}^{4}$$

Thus losses are defined by positive numbers, and gains by negative numbers.

It will be noted that speech pressures, rather than speech powers, have been chosen as references; this has been done for two reasons. In the first place, speech power is very difficult, whereas speech pressure is relatively easy, to measure. Secondly, the ear appears to be a pressure sensitive rather than a power absorbing device, in the same sense that the input circuit of a vacuum tube is considered to be voltage actuated rather than power absorbing. When a telephone receiver is held close to the ear, the acoustic impedance presented to it is largely reactive. The real part of the acoustic impedance is largely due to the leakage path to the outside air between the receiver cap and the ear. Consequently the acoustic power is dissipated, for the most part, in this leakage path. Hence, whereas power rating would provide a satisfactory measure of the merit of a receiver as a loudspeaker, it seems more reasonable to use pressure in rating its effectiveness as a transducer when held close to the ear.

LOUDNESS RATING OF TRANSMITTING COMPONENTS

When figures of merit are to be assigned to the components of a telephone connection apart from the over-all rating described in the previous section, it is convenient first to divide the connection into two basic categories, one of which includes the transmitting transducer and the other the receiving transducer. In this section we consider components which include the transmitting transducer, while those which include the receiving transducer are treated in the following section.

For engineering purposes the transmitting component of greatest interest is probably the transmitting telephone set itself. A transmitting telephone set has a transmission rating of zero when the voltage (in volts) it delivers across a 900-ohm resistive load is equal to the pressure (in millibars) at the talker's lips. As in all ratings involving transmitters, the measurement of talking pressure is a free field one with the transmitter

[4] Since the ratios in this equation and those which follow involve either pressures across dissimilar acoustic impedances or an acoustic pressure and an electric voltage, the resulting ratings should not be expressed in decibels. However, the equations yield db-like quantities which may be measured on meters equipped with scales calibrated in decibels which will give numerically accurate results. The use of logarithmic units is highly desirable since this makes possible the direct addition of component ratings to give the over-all rating of a particular connection.

removed from in front of the sound source, while the output measurement is made with the transmitter in the modal position with respect to the sound source. Thus the rating, R_T, of any transmitting set is:

$$R_T = -20 \log_{10} \frac{V_T}{S_T}^4$$

where V_T is the voltage across the load and S_T is the talking pressure in millibars. It may be noted that the same numerical ratings will be obtained if the voltage is expressed in millivolts and the pressure in microbars.

Since the transmitting efficiencies of commercial telephone sets vary with the direct current in the transmitter, it is necessary to specify an appropriate value of line current or voltage representative of typical operation, when expressing the transmission rating of a transmitting telephone set. For present commercial sets with carbon-type transmitters the value of line current specified is 100 milliamperes, which is appropriate because it represents nearly the maximum dc obtainable on short loops. Any departure from 100 milliamperes in a commercial telephone connection using present sets causes a current supply loss (or gain), which is added algebraically to the transmission rating of the transmitting telephone set to obtain the transmitting conversion loss. The sign of the transmitting current supply loss will usually be positive for currents less than 100 ma, and negative (a gain) for currents greater than 100 ma.

Transmitting loop losses can be defined and measured in the same way as the rating of a transmitting telephone set if the 900-ohm resistive load is located at the output (central office end) of the loop instead of at the output of the transmitting telephone set. In this case the rating of the

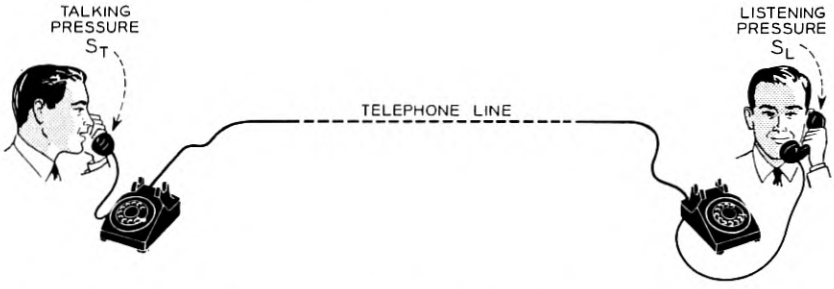


Fig. 1 — Loudness rating of an over-all telephone connection $= -20 \log_{10} S_L/S_T$. Rating is zero when $S_L = S_T$.

loop is equal to the sum of the rating of the transmitting telephone set, the transmitting current supply loss, and the insertion loss of the subscriber line between the impedance of the telephone set and the 900-ohm termination. The last is called the subscriber line loss.

LOUDNESS RATING OF RECEIVING COMPONENTS

As in the case of rating transmitting components the item of greatest engineering interest in the category containing receiving transducers is the receiving telephone set itself. To be consistent with the foregoing definitions for transmitting components and for over-all connections, its rating is defined as follows. A receiving telephone set has a loudness transmission rating of zero when the output acoustic speech pressure (in millibars) delivered by the receiver to the ear of a listener, is one half the open circuit voltage (in volts) of a 900-ohm resistive source which energizes the set as a load. The transmission rating, R_L, of any telephone set when receiving, then is

$$R_L = -20 \log_{10} \frac{S_L{}^4}{V_W/2}$$

where S_L is the listening pressure in millibars and V_W is the open circuit voltage of the 900-ohm source. As before, pressure in microbars and voltage in millivolts will give the same numerical ratings.

The same appropriate value of line current or voltage that is specified for transmitting ratings is also specified when expressing the transmission rating of a receiving telephone set. For present sets, as stated, the value is 100 milliamperes line current. Departures from this value cause a receiving current supply loss in telephone sets with current controlled equalizers. For currents less than 100 milliamperes the sign of this loss is usually negative because of the increased circuit efficiency on longer loops. The sum of the rating of the receiving telephone set and the receiving current supply loss is the receiving conversion loss. In present telephone sets without equalizers the receiving current supply loss is zero, so the receiving conversion loss is the same as the receiving rating of the set.

The foregoing equation also expresses the transmission rating of a receiving loop provided the 900-ohm source is applied to the input (central office end) of the loop instead of to the receiving telephone set. The rating of the loop, so obtained, is equal to the sum of the rating of the receiving telephone set, the receiving current supply loss, and a loss called the receiving subscriber line loss which is equal to the insertion loss of the

subscriber line between 900 ohms and the impedance of the telephone set. If like facilities and instrumentalities are used for both transmitting and receiving loops, the transmitting and receiving subscriber line losses will be the same, although the loop losses will not be the same because of the different set ratings.

ADDITIVE PROPERTY OF COMPONENT RATINGS[5]

It was pointed out earlier that consistency of the definitions of the component ratings may be demonstrated by showing under what conditions the sum of the component ratings is equal to the over-all rating. The simplest telephone circuit that can be visualized for this purpose is illustrated in Fig. 2.

Here two commercial type telephone sets, each supplied with 100 milliamperes dc, are connected to each other by way of an imaginary amplifier inserted for the purpose of providing a pure resistive impedance of the right magnitude. This amplifier has input and output impedances each equal to 900 ohms pure resistance, and an image gain of zero decibels. It will be noticed that the transmitting telephone set has a 900-ohm resistive load impedance, so the ratio of the voltage V_T, across the input of the amplifier, to the talking pressure, S_T, is a measure of the transmitting rating of the telephone set by definition. Also, the output of the amplifier is a 900-ohm resistive voltage source for the receiving telephone set. If the open circuit output voltage of the amplifier is V_W, then the ratio of the listening pressure, S_L, to $V_W/2$ (shown by the dashed arrows in Fig. 2) by definition measures the receiving rating of the telephone set. Furthermore it can be shown by network theory that the open circuit output voltage of an amplifier of zero db image gain is twice the voltage across its input, provided its output and input impedances are equal. In other words, in this circuit,

$$V_W/2 = V_T .$$

Since this is so, the over-all rating of this telephone connection, measured by the ratio of listening to talking pressure, is equal to the sum of the transmitting and receiving ratings of the respective telephone sets.[6]

It is apparent from this discussion that if the two telephone sets are

[5] While it is recognized that the additive property of component ratings can be only approximate if a wide variety of frequency spectra is taken into account, nevertheless this is of little practical concern in the case of spectra usually encountered in the modern telephone plant. In this paper it is convenient for the sake of simplicity to disregard the approximation.

[6] A further note on the effect of frequency characteristics of components on the additive property of component ratings is included in Appendix A.

connected to each other back to back, without including the hypothetical amplifier of Fig. 2, the sum of their component ratings must be augmented by a reflection loss in order to equal the over-all rating. This comes about because the choice of definitions for transmitting and receiving ratings was dictated by the practical consideration that the more difficult transmission conditions occur when lines or apparatus are interposed between the sets. Thus it is advantageous to have the basic ratings more appropriate to these actual conditions.

By the definitions given here, the sum of two loop losses will also equal the over-all rating if the loops are connected to each other through the hypothetical amplifier. The majority of all actual telephone connections involves two loops connected to each other by some form or combination of trunks, instead of by the amplifier. The inclusion of trunks in the connection augments the two loop losses by the attenuation, reflection and interaction losses associated with the trunks. While there is considerable variation in the impedances of facilities making up loops and trunks, the resulting reflection and interaction losses can be minimized, and in many cases may be neglected altogether, by the choice of a compromise impedance for the rating definitions, which is reasonably representative of the entire distribution of loop and trunk impedances. These considerations dictated the particular choice of 900 ohms resistance as the terminating and source impedances in the definitions of transmitting and receiving component ratings.

One other aspect of the selection of definitions of transmitting and receiving ratings will now be discussed. We mentioned earlier that it seemed more logical to base the rating system on ratios of pressure and voltage than on power ratios because the human ear appears to be a pressure sensitive device. As a result of this decision it becomes apparent on reflection that the *numerical* apportionment of an over-all loss between

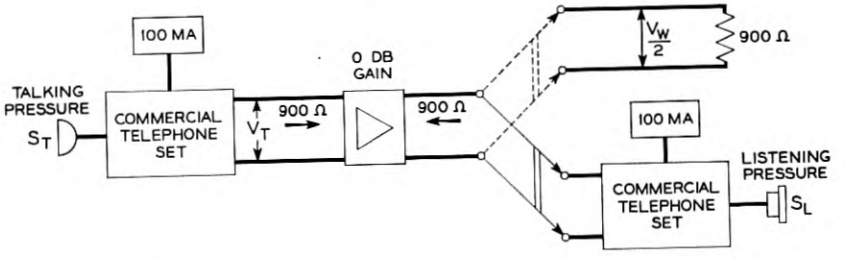


Fig. 2 — Additive property of transmission ratings. Transmitting rating: $R_T = -20 \log_{10} V_T/S_T$. Receiving rating: $R_L = -20 \log_{10} (S_L/(V_W/2))$. Since $V_w/2 = V_T$, over-all rating, $R_0 = -20 \log_{10} S_L/S_T = R_T + R_L$.

transmitting and receiving components is an arbitrary choice. The range of numbers involved can be controlled by the units in which the voltages and pressures are expressed, or by inclusion of an arbitrary numerical factor by which the voltage-pressure ratios are multiplied, without affecting the additive property of the component ratings. It is desirable to obtain an appropriate range of numbers by means of the simplest possible definitions of the transmitting and receiving ratings, if possible avoiding the use of arbitrary multiplying factors. This objective has been attained without multiplying factors by specifying that the voltages be expressed in volts and the pressures in millibars, or, which comes to the same thing, millivolts and microbars.

The range of numbers is appropriate because of the choice of units in the definitions. The transmitting ratings of present commercial sets, transmitting conversion losses and loop losses will all have negative signs, while the corresponding receiving ratings will all be positive. This is consistent with the fact that carbon transmitters produce gain and electromagnetic receivers loss as transducers between acoustic and electric energy. Even if more efficient receivers of this type are developed in the future, the definition of receiving ratings is such that the numerical values can never become negative without the aid of a gain producing device such as an amplifier.

LINE LOSSES

The line is here considered to be that portion of an over-all telephone connection between the line terminals of a transmitting telephone set and those of a receiving telephone set. In the simple circuit of Fig. 2 the "line" consists solely of the hypothetical amplifier, and the line loss, considered as a separate entity, is zero. In this case the over-all rating, as already stated, is equal to the sum of the transmitting and receiving ratings of the telephone sets. More generally, if the dc supplies for present sets are other than 100 milliamperes, the over-all rating is equal to the sum of the transmitting and receiving *conversion* losses, since conversion loss was defined as the sum of set rating and current supply loss.

If, now, the hypothetical amplifier is replaced by an actual line, consisting perhaps of two subscriber lines, battery supplies and a trunk, we are interested in the amount of loss which must be added to the conversion losses in order to equal the new over-all loss. This added amount defines the line loss, considered as an entity by itself. It is composed of reflection losses (involving the differences between reflection losses of the sets against the impedance of the line and the sets against 900 ohms),

attenuation losses and interaction losses. As a practical matter, however, line loss is defined as the difference between the over-all rating of the telephone connection and the sum of the transmitting and receiving conversion losses.

The breakdown of line losses into components suitable for engineering uses is beyond the scope of this paper. Little departure from established practice in this matter is made necessary by the new transmission rating system. The general principles which have been discussed may be helpful in taking care of such departures as appear. As a final example, loop losses have been defined here as the sum of conversion losses and subscriber line losses. If loop losses are to be used for engineering purposes, it is then possible to define "connecting circuit loss," caused by a trunk or combination of trunks, as the difference between the over-all rating of the telephone connection and the sum of the two loop losses. Thus the line loss is equal to the sum of two subscriber line losses and the connecting circuit loss.

ELECTRO-ACOUSTIC MEASURING SYSTEM

It was stated earlier that one of the advantages in going to a loudness basis in the new rating plan is that it permits objective transmission measurement, whereas the subjective factors included in the effective loss system of rating make this impossible. It was pointed out that loudness itself is a subjective matter, but that objective measurements can be made which usefully approximate, although they do not duplicate, loudness determinations. While other systems are being investigated, the assembly of apparatus which has been used so far in the laboratory for this purpose is described in this section. It is called the electro-acoustic transmission measuring system. A functional diagram of its parts is shown in Fig. 3.

It is convenient for descriptive purposes to divide the electro-acoustic measuring system into three parts:

(1) The sound source, which takes the place of a human talker at the talking end of a telephone connection, as shown in Fig. 3(a).

(2) The indicating meter for acoustic pressure or electrical voltage. This is a substitute for a human listener at the receiving end of a telephone connection, and is shown in Fig. 3(b).

(3) The basic telephone connection, an intermediate link between sound source and indicating meter, which is provided with stable transducers as a substitute for commercial telephone connections for purposes which will be described further on. This appears in Fig. 3(c).

The sound source consists essentially of an electric generator of complex voice frequency waves and an artificial mouth.[7] The electrical output of the tone generator is shaped by a network to have the frequency spectrum of human speech. An equalizer is included in the circuit to correct the sound field from the artificial mouth so that it will be flat for a flat electrical input. An amplifier with variable gain permits adjustment of the acoustic level at the output of the artificial mouth to values comparable to the levels of human speech. One form of complex tone generator which has been used is a warbler oscillator in which the frequency of oscillation is varied linearly with time from 300 cps to 3,300 cps and back to 300 cps, six times per second. The sharp changes in amplitude caused by modulation products, and the frequency range, result in an acoustic output from the artificial mouth which has some of the characteristics of human speech, insofar as activation of commercial carbon type transmitters is concerned. In addition to the warbler oscillator, other types of tone source are being investigated.

The indicating meter shown in (b) of Fig. 3 is adaptable either to measurements of the acoustic pressure of real or simulated speech at the two ends of a telephone connection, or to speech voltage measurements at an intermediate point. For pressure measurements a condenser microphone is used as a transducer. Voltage measurements are made by means of a high impedance bridging connection. A high-pass filter is employed to reduce components of noise, particularly 60-cycle hum and its immediate harmonics, which are below the frequency range normally passed by telephone circuits. The meter circuit also includes a loudness shaping network which weights indications so that the difference between two measurements represents, approximately at least, the loudness difference perceived by the human ear.[8]

The basic telephone connection, parts of which may be used with portions of a commercial telephone connection to make up an over-all connection, is shown in Fig. 3(c). This is a two-way four-wire connection which makes use of electromagnetic, rather than carbon type, transmitters in order to avoid the instabilities of the latter type. However, shaping networks are employed in the transmitting portions of the two converting sets to provide the frequency characteristics of modern carbon type transmitters. The characteristic is roughly a compromise between those of the F1 and T1 transmitters. The line impedances of both trans-

[7] A. H. Inglis, C. H. G. Gray and R. T. Jenkins, A Voice and Ear for Telephone Measurements, B. S. T. J., **11**, p. 293, April, 1932.

[8] H. F. Hopkins and N. R. Stryker, Proposed Loudness-Efficiency Rating for Loudspeakers and the Determination of System Power Requirements for Enclosures, I. R. E. Proc., **36**, p. 315, March, 1948.

mitting and receiving portions of the converting sets are 900 ohms pure resistance, matching the 900-ohm distortionless connecting lines without reflection. The gain of amplifiers in the transmitting and receiving portions of the sets may be adjusted so that the sets have convenient transmitting and receiving ratings in terms of the new rating system. An additional amplifier in each converting set provides an adjustable sidetone path so that the connection will not sound dead to human talkers and listeners.

METHODS OF USE

Examples of various methods of use of the electro-acoustic measuring system are given in Figs. 4, 5, 6 and 7. Since these are laboratory meas-

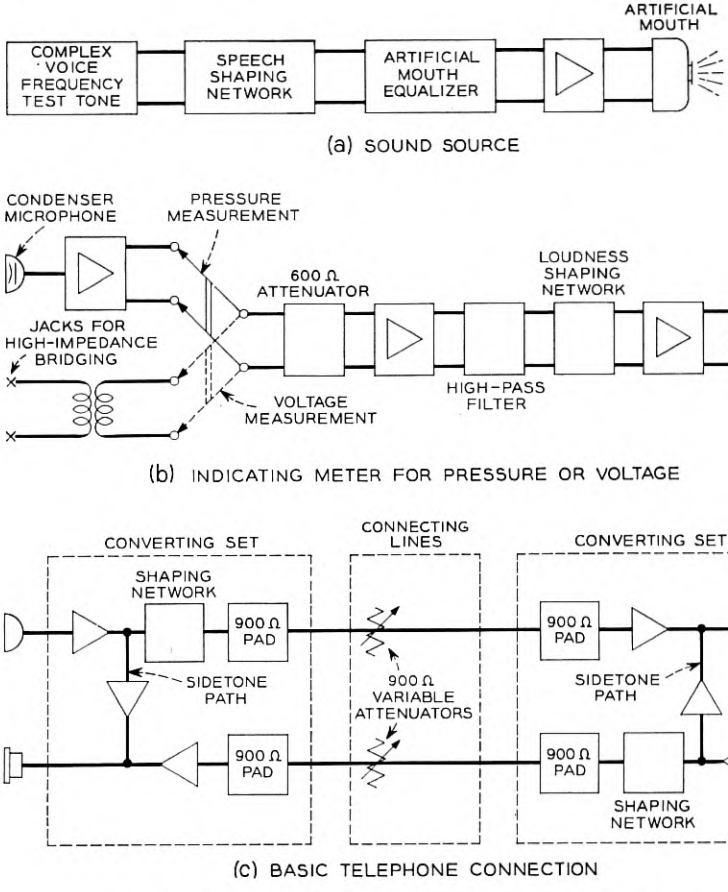


Fig. 3 — Electro-acoustic transmission measuring system.

urements, line facilities such as loops and trunks are simulated by equivalent networks. Thus both ends of telephone connections, which might be far apart in the actual case, are readily available for measurements in the laboratory.

The first example, Fig. 4, illustrates the use of the sound source and indicating meter to measure the over-all rating of a complete telephone connection. During measurement of the talking pressure, S_T, the telephone handset is removed and the condenser microphone placed at a carefully gauged position with respect to the guard ring of the artificial mouth. When this measurement has been made the handset is replaced, with the transmitter positioned by template. Different templates are necessary for F-type and G-type handsets so that each will be in the modal position with respect to the source of sound. The condenser microphone is then removed from its artificial mouth position and placed in the 6-cc coupler at the listening end of the connection. The telephone receiver is also clamped to the coupler which furnishes an acoustic termination similar in impedance to that provided by the human ear when the receiver is held tightly against it.[9] Prior to the measurement of listening pressure, as in all measurements involving carbon transmitters, the carbon must be conditioned. This is a process of physical agitation in a prescribed manner, and a momentary acoustical overload, which eliminates packing of the carbon which might otherwise occur when the handset is held rigidly in a fixture. It places the carbon in a configuration similar to that experienced under normal talking conditions. After conditioning the measurement of listening pressure is made.

The use of the sound source and indicating meter for rating a transmitting telephone set is shown in Fig. 5. Here the output measurement is one of voltage rather than pressure. For this the indicating meter is bridged to the circuit by means of a high impedance connection.

Receiving rating of a telephone set is illustrated in Fig. 6 which shows measurements of the source voltage and listening pressure. The frequency characteristic of the source voltage includes not only speech weighting but also the response characteristic of carbon transmitters. This is necessary in order that the component ratings may add up to the over-all rating, as indicated in Appendix A. A convenient way of supplying the desired frequency characteristic is to use the artificial mouth and the transmitting converting set of the basic telephone connection. The output from the 900-ohm connecting line is a voltage of proper frequency characteristic from a 900-ohm source, as required for

[9] E. E. Mott and R. C. Miner, The Ring Armature Telephone Receiver, B. S. T. J., **30**, p. 110, Jan., 1951. Fig. 4 in this reference shows the similarity in receiver response when measured in a 6-cc coupler and when measured against the human ear.

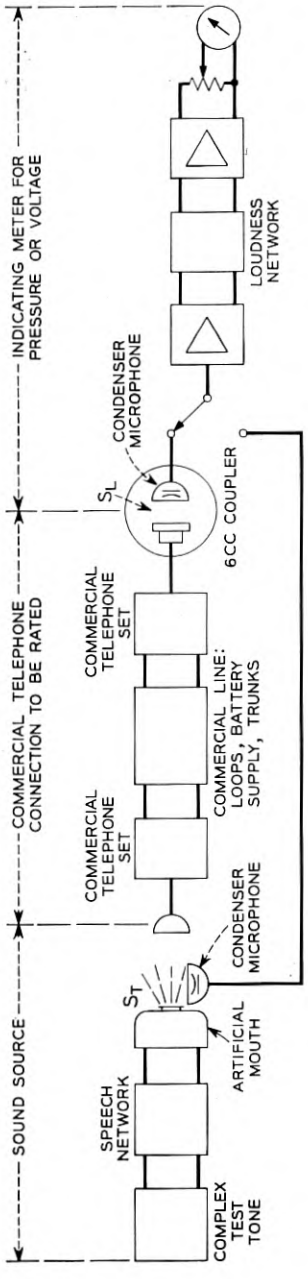


Fig. 4 — Measurement of over-all rating of a telephone connection.

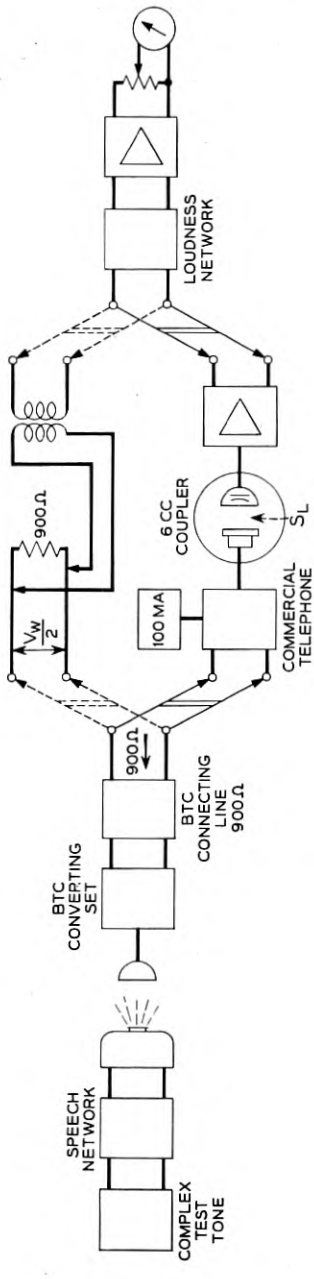


Fig. 6 — Measurement of receiving rating of a telephone set.

the receiving rating of commercial telephone sets. The measurements of source voltage and listening pressure are made in the same manner as in the previous examples. No conditioning is necessary prior to these measurements because the transmitter of the converting set is electromagnetic instead of carbon type.

Another use of the basic telephone connection is to obtain basic data for subjective rating by means of comparison tests. Subjective ratings require human talkers and listeners. An example of this use is shown by Fig. 7 in which a transmitting telephone set is rated subjectively in order to check data obtained by objective measurement. In this test the talker speaks alternately into the transmitter of the set to be rated and into that of the converting set of the basic telephone connection. The listener listens at the receiver of the basic telephone connection to speech

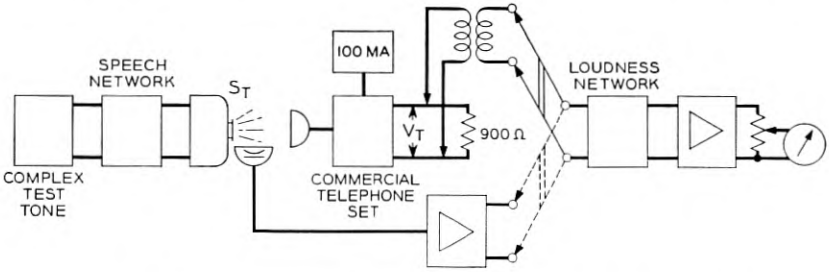


Fig. 5 — Measurement of transmitting rating of a telephone set.

from both sources. He adjusts the attenuator (B) in the basic telephone connection until he judges the two to be equal. The difference between the settings of the two attenuators is then a measure of the rating of the commercial telephone set relative to that of the converting set in the basic telephone connection. The basis of equality in the observer's judgment must be specified; loudness may be specified, or some other subjective factor such as general satisfaction. Subjective tests of this kind require a number of different listeners, and often a number of different talkers, to yield results that are useful. Other arrangements of the basic telephone connection can be used for similar comparison tests of other commercial telephone components.

APPARATUS FOR FIELD TESTING

As already stated, the transmission measuring system described in the previous sections was assembled to obtain laboratory data which will be

of use to transmission engineers in the field. The data so obtained are in terms of the transmission rating system discussed in this paper. It requires no great stretch of the imagination to visualize the development of apparatus functionally similar to some of that which has been described, but much smaller, lighter in weight and far less complex, which might be used by field personnel. Elements of such a measuring system would be located in central offices. Other elements might be carried to subscribers' locations in the normal complement of tools used by installation or maintenance forces. This field transmission measuring apparatus would be used to make field transmission surveys. It might also

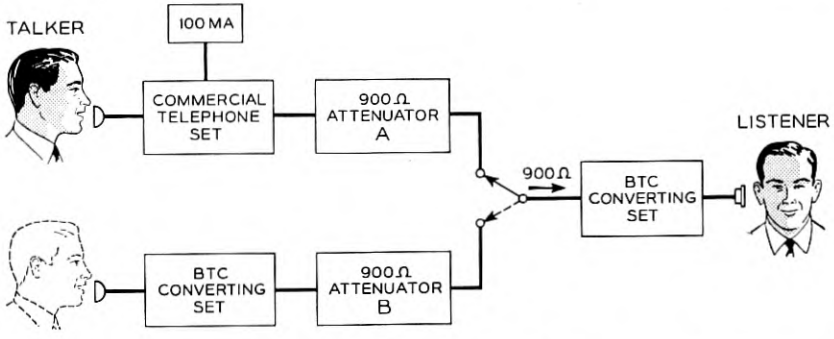


Fig. 7 — Use of basic telephone connection in subjective rating of a transmitting telephone set on a comparison basis.

be used for maintenance purposes: values of electro-acoustic transmission loss measured in the field under circumstances of suspected trouble could be compared with transmission design data. The possibilities of such uses are wide in scope. These possibilities were in mind when the loudness rating plan was devised.

SUMMARY

Some of the disadvantages of the effective loss system for rating transmission were mentioned in the introduction. It is confidently expected that the revised rating plan presented here is free from these disadvantages and has merits of its own. The basis for this confidence lies in the following advantages of the plan, which are briefly summarized.

References are dispensed with in the loudness rating plan. Instead of references, ratings are in terms of absolute physical quantities which can be readily understood, and the ratings, therefore, can be visualized. This

should lead to a far wider understanding of transmission rating, and of the meaning of rating numbers.

Ratings for the transmitting and receiving components of telephone connections have additive property and, for modern station instrumentalities, result in a range of rating numbers which is appropriate.

Loudness ratings may be measured objectively, instead of being determined by time consuming subjective methods.

The rating plan permits development of light weight field transmission measuring apparatus for field survey and maintenance purposes.

In special cases, where loudness rating is inadequate, subjective ratings may be applied by the addition of transmission penalites to loudness ratings.

At the present time engineers at Bell Telephone Laboratories are engaged in an extensive program of measurements so that the new rating plan may be implemented as soon as possible.

ACKNOWLEDGMENT

The authors are indebted to their many colleagues past and present in the Bell System who have been associated at one time or another with the problem of transmission rating. This problem, looked at over the years, is one in which the requirements and needs are continuingly variable. To be successful in the solution of such a problem, a new rating plan must include the best features of all past plans with such new features as may be necessary for current needs. It is patently impossible to give individual credit where it is due. However, the authors wish to acknowledge the contributions made by N. R. Stryker to the instrumentation used to date.

Appendix A

THE EFFECT OF FREQUENCY CHARACTERISTICS ON THE ADDITIVE PROPERTY OF COMPONENT RATINGS

It is recognized that, strictly speaking, there is no known way to define component losses in such a way that all of them are independent of the remainder of the circuit with which they are used, and at the same time that their individual ratings add up to the over-all rating under all conditions. Consider as an extreme example an over-all connection in which the transmitting component produces frequencies only in the band from zero to 1,500 cps, and the receiving component is responsive to frequencies in the band from 1,500 to 3,000 cps. If the transmitting

circuit alone were measured in a wide band measuring circuit it would be assigned a rating of, say, X. Likewise if the receiving circuit by itself were measured by a wide band measuring circuit it might be assigned a rating of Y. The two added together would provide an over-all rating of X + Y; yet obviously the actual over-all loss would be infinite. If, however, the response-frequency characteristic of the transmitting circuit were included in the source voltage for the receiving circuit measurements, the receiving rating would no longer be Y, but would be an infinite loss. In this case the receiving rating would not be independent of the rest of the circuit, but the sum of the transmitting and receiving ratings would be equal to the over-all rating.

An alternative would be to apply the response-frequency characteristic of the *receiving* circuit to the measuring circuit when measuring the output of the *transmitting* circuit. In this case it would be the transmitting circuit which shows an infinite loss, while the receiving rating would be finite; the transmitting rating would not be independent of the circuit, but the two would add up to the over-all rating.

In the example cited, which is of course an absurd extreme, one is left with the dilemma: which of the two circuits, transmitting or receiving, should be the one to be rated as an infinite loss? There is no ready answer to this question. Fortunately, in more practical cases, the choice becomes an academic one because, first, the telephone plant is designed so that its components work together rather than in opposition to each other, and secondly the response of receiving circuits is in general relatively flat with frequency. The first of these reasons means that there would be little difference in the component ratings if either alternative were adopted. The second reason permits a method of measurement under the second alternative which still allows the component ratings to add up to the over-all rating.

In the rating plan described in this paper the first of the two alternatives mentioned above was chosen: namely, that transmitting ratings are determined by measurements independent of the remainder of the circuit, while receiving ratings are determined by measurements in which the source voltage includes the frequency-response characteristic of modern carbon transmitters. Three reasons dictate this choice:

(1) It permits simple definitions of component transmitting and receiving ratings.

(2) The input and output measurements for the component ratings are conveniently made.

(3) Greater generality is provided because the additive property of component ratings does not depend on flatness of the receiver characteristic.

Protection of Service in the TD-2 Radio Relay System by Automatic Channel Switching

By I. WELBER, H. W. EVANS, and G. A. PULLIS

(Manuscript received December 3, 1954)

The TD-2 radio relay system spans the continent and provides for transmission of hundreds of telephone conversations or several television programs. The effect of even an occasional failure would be so great that means have been provided to restore service quickly and automatically when a failure occurs. The operation and circuits of this automatic channel protection switching are described.

INTRODUCTION

When communication systems fail, provision must be made to restore service as rapidly as possible. In systems providing only a few telephone circuits, restoral may be by manual rearrangement of the circuits affected so that they are made good on stand-by or alternate facilities. But in the case of systems carrying a large number of telephone channels or network television, the impact of a system failure is so great that automatic arrangements must be provided to restore service much faster than is possible by manual operations. This need led to the development of the automatic protection switching system for the TD-2 radio relay system,[1] which has in recent years come to provide a large portion of the Bell System's long-haul telephone and television facilities.

The TD-2 system provides six two-way radio channels, any one of which will carry several hundred telephone circuits or one television signal. It is used generally as a long-haul backbone route system, principally between major centers of population. Repeater stations are provided at intervals of about thirty miles, depending on considerations of terrain, and the number of repeaters between the terminal points is a function of course of their geographical separation. Frequency modulation is used, and the repeaters do not reduce the signal to its original amplitude modulation form, so that the base-band is not normally available at repeaters intermediate to the terminal points.

One of the six two-way radio channels is reserved as a protection channel for use when a regular channel used in telephone or television business has been interrupted. The automatic protection switching system consists, in general terms, of devices which detect an interruption or degradation of transmission of each regular channel, and devices which substitute the protection channel for the regular channel in trouble. The latter devices comprise both switches and circuits to control them.

Some TD-2 systems are so long that it is desirable to divide them into shorter sections called Switching Sections. The protection channel is subject of course to the same vicissitudes that interrupt the regular channels. These vicissitudes occur on a per repeater section basis, so that the greater the number of repeater sections in a switching section, the greater the time that the protection channel will not be available, and also the greater the total time that the regular channels will require the protection channel. It will be shown that the probability of interruption of a regular channel is proportional to the number of repeaters in each switching section. The number of repeaters per switching section represents an economic balance between the degree of reliability desired and the cost of the automatic switching equipment. In practice, the average number of repeaters per switching section is 10, with a range of from 5 to 15 except for certain special cases where even fewer than 5 repeaters are protected.

Since switching arrangements must be provided at the ends of the switching sections, which may be at repeaters other than terminals, it is necessary to examine the condition of the radio channels at intermediate frequency rather than at baseband. A channel switching initiator is associated with each radio channel at the receiving end of each switching section, and comprises a frequency modulation receiver and a noise detector. It will detect the presence of abnormally high noise near 8.5 mc, well above the frequencies usually transmitted over the system, and the absence of carrier. Either condition is symptomatic of system failure or degradation.

Each switching section is a separate entity, in that information is not exchanged between adjacent switching sections. Thus if a failure occurs in the first of several switching sections, high noise may be detected by the channel initiators in all the switching sections, even though a switch to the protection channel is needed in only the first section. Spurious switches are prevented by switching control arrangements. When a channel initiator indicates a possible failure, the regular channel with which it is associated is bridged at the transmitting end of the switching

section to the protection channel. If transmission is found to be deficient on both the regular and protection channels, as would be the case for troubles in a preceding section, no switch results. This feature also prevents switches from a marginal though usable regular channel to a totally interrupted protection channel.

SERVICE INTERRUPTIONS

Radio systems are subject to two types of interruptions: equipment failures, which will usually completely interrupt transmission until repairs can be made by men dispatched to the point of failure, and radio fades. The duration of fades is short compared to that of equipment failures but the fades occur so frequently during certain periods that they are objectionable.

Equipment Failures

Even though the individual tubes and components that make up the equipment of the TD-2 system have very creditable life figures, they do sometimes fail, and the impact of their failure on system performance is serious because of the enormous number of tubes and components involved. Thus between Los Angeles and New York in each direction of transmission there are more than 450 416-type Western Electric vacuum tubes used as microwave amplifiers or modulators and more than 1,100 tubes of other types directly in the transmission path, plus over 1,000 tubes used in such auxiliary functions as the generation of micro wave frequencies. Failure of any one tube or any of its associated components will cause an interruption to transmission. Every possible precaution is taken to avoid these failures, and considerable improvements have been made in repeater station reliability since the TD-2 system was first placed in service.

Radio Fading

Microwaves used for radio relay systems propagate through the atmosphere, which is a dielectric whose characteristics vary with pressure (altitude), moisture content, temperature, etc. These factors may vary from point to point along the path to such an extent that the transmission medium is so nonhomogeneous that several discrete transmission paths exist between a radio transmitter and a radio receiver.[2] These paths may differ in length so that at some frequencies signals may arrive via two paths with such phase relationship that one signal cancels the

other, resulting in a radio fade.* These fades are usually very short in duration compared to equipment failure interruptions because the dielectric character of the transmission path is continually changing with motion of the air. Since the existence of a fade is predicated on a natural coincidence of factors involving frequency, position of the antennas relative to one another and the micrometeorology of the path, it is apparent that fades are not likely to occur at two frequencies at the same time, or on two paths at the same time. Thus the effect of radio fading can usually be counteracted by use of either space diversity, which provides two transmission paths and selects the better, or frequency diversity, which provides two transmitting frequencies and selects the better. Frequency diversity was chosen for the TD-2 automatic switching system because it also provides more thorough protection against equipment failure than is possible with space diversity.

Radio fading in most sections of the country is most apt to occur in summer months when the moisture content of the air is higher. Also during the summer months fading is most apt to occur during the night hours when there is less air motion than in the day. Fortunately the worst fading occurs during early morning hours, say 3 to 5 A.M., when television and telephone demand has been low. The fading observed on a single path in Ohio during a typical summer night is shown in Fig. 1. The net effect of a fade is an increase in telephone or television circuit noise, as shown in Fig. 2 for the telephone case. If the fade is so deep that the carrier-to-noise ratio at the repeater input approaches unity, the signal drops off as well, but this is beyond the point where noise is considered intolerable.

Fading measurements have been made on many paths in many parts of the country for considerable periods of time, and also several continuous observations of telephone channel noise have been carried on, so that a considerable body of data was available on which to base predictions of fading performance.

The maximum rate of change in path loss with time during a fade is also of interest, since this is one of the factors that determines the time within which the protection system must operate. Several hundred deep-fades were examined, and the average maximum rate of change in path loss was found to be about 10 db per second. It is believed that rates as high as 100 db per second occur very seldom. Because the switching system takes approximately $\frac{1}{20}$ second to operate, the path loss at the

* There are other types of fades, as described in Reference 2, which have other causes. Multipath transmission causes almost all of the fading observed on TD-2 systems. "Earth-bulging" fades have been observed, but very rarely.

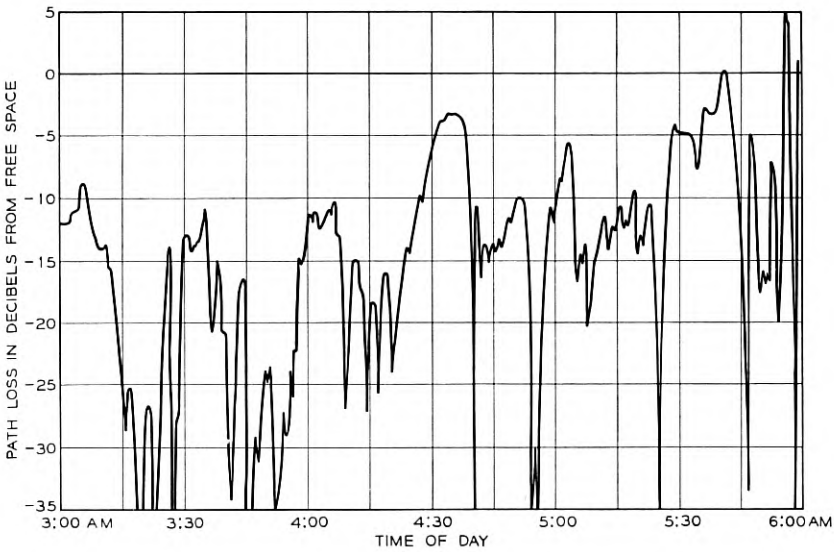


Fig. 1 — Heavy summer fading, Bryan-Wauseon, Ohio.

actual instant of operation may be 0.5 db greater than at the instant the channel initiator called for a switch, and for some fast-moving fades this figure may be as much as 5 db.

Expected Improvement in Reliability

The objective of the automatic protection switching system is, of course, to make the TD-2 system as reliable as possible, consistent with frequency availability and economic considerations. The frequency space available in the 4,000-mc common carrier band permits operation of six broadband channels on a route, and if the frequency space is to be used most efficiently only one channel can be used for protection purposes. Perfect reliability cannot be achieved because (a) the protection channel is subject to the same causes of failure as are the regular channels and (b) occasionally two regular channels fail at the same time, and the single protection channel can restore only one of the failed channels.

Obviously the greater the number of regular channels protected by a single protection channel, the higher the probability that simultaneous failures will occur on two regular channels. Also, the greater the number of repeaters in a switching section, the greater the number of demands on the protection channel and the greater the probability that the protec-

tion channel itself will fail. It is economically desirable that switching sections be long, since switching equipment is expensive. The number of points at which switching equipment must be provided depends on the degree of improvement needed to increase the reliability to an acceptable point.

It is convenient to discuss the degree of improvement in reliability separately for equipment failures and for fading. The probability of interruption with automatic switching of a one-way channel because of equipment failure is shown in Appendix I to be:

$$P = R \frac{(N+1)}{2} np^2$$

where p = the probability that any one-way repeater will be in a failed condition,

n = the number of repeaters per one-way channel in each switching section,

N = the numbers of regular channels, and

R = the total number of repeaters per one-way channel.

From data on incidence and duration of equipment failures an average figure for p is

$$p = 0.05 \text{ percent} = 0.0005$$

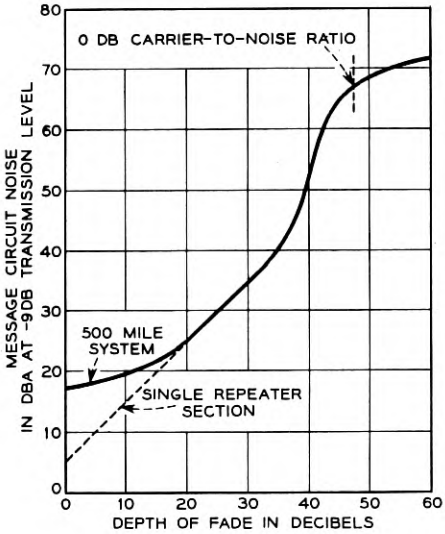


Fig. 2 — Fluctuation noise in the top channel of a 480-channel message system due to a fade in one repeater section.

Typical values for n and N are

$$n = 10$$

$$N = 5$$

Reliability figures are frequently quoted on a 4000-mile basis, since this is representative of the longest circuits the Bell System provides. This corresponds to a total of 133 repeaters, assuming 30-mile spacing, so

$$P = 0.001 = 0.1 \text{ per cent}$$

This is less than two minutes per day.

Without automatic protection switching, the probability of interruption of a one-way 4,000-mile (133 repeaters) TD-2 channel would be 6.65 per cent or 96 minutes per day, which is obviously excessive. Thus, the outage time due to equipment failure with no protection whatsoever is 66 times the outage time with automatic protection switching. This figure, interestingly enough, is independent of the length of the system.

Of course the automatic system replaces manual switching rather than no protection at all, and it is difficult to estimate the effectiveness of manual switching, so that the degree of improvement over manual switching cannot be stated.

The improvement in reliability associated with atmospheric fading is computed in the following manner. Let us consider a typical switching section of ten repeater sections, and let us assume that the fading over this route is equivalent to that observed experimentally on the New York-Chicago TD-2 system between midnight and 8 A.M. (heaviest fading hours) in August (heaviest fading month) 1950 (typical fading year). The noise distribution for the unprotected system is shown in Fig. 3, designated, "No Switching." The noise in the top channel of a 480-channel telephone system would exceed 40 dba at the -9 db transmission level point 0.11 per cent of the time. In a 4,000-mile system made up of thirteen such switching sections the noise without automatic protection switching would exceed 40 dba 1.43 per cent of the time. This is excessive.

Now let us assume that the ten repeater switching system is equipped with an automatic switching section arranged to switch to a protection channel when the message circuit noise exceeds 40 dba. As in the case of equipment failures, the protection channel is not always available when needed. It might have failed or might be busy making good another regular channel which has failed. Also, the protection channel at the instant when it is needed might be experiencing a fade even deeper than the regular channel with which we are concerned, and in

this case the switching system must be able to recognize that no switch should be made. This probability was evaluated by R. L. Kaylor from experimental studies of the loss versus frequency of a large number of fades.[3] As would be expected, there is a greater probability that a fade will affect two channels immediately adjacent in frequency than two channels separated in frequency. For this reason the protection channel is wherever possible at the extreme end of the band (usually the low-frequency end, since channels are installed in sequence, and a protection channel is needed no matter how many regular channels are installed).

Summarizing these factors for the switching section of ten repeaters and five regular channels, the protection channel availability is given in Table I.

Since the protection channel is unavailable 9 per cent of the time, the fraction of the time that the telephone circuit noise exceeds 40 dba at the −9 db transmission level point will be reduced from 0.11 per cent of the time when unprotected to 9 per cent of 0.11 per cent, or 0.01 per cent of

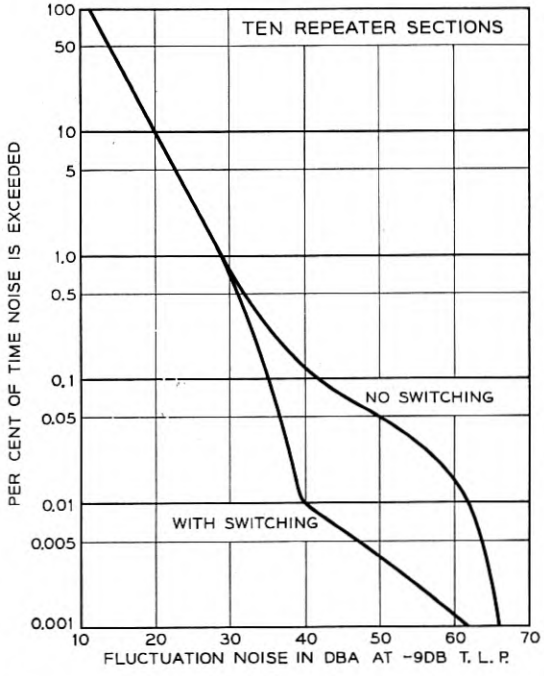


Fig. 3 — Fluctuation noise in the top channel of a 480-channel message system during heavy summer fading, with and without automatic protection switching.

TABLE I — PER CENT OF TIME PROTECTION CHANNEL IS NOT AVAILABLE
Hypothetical Six-Channel Ten Repeater TD-2 System
Heavy Summer Fading

Condition	Per cent Time Unavailable
Fading: protection channel worse than regular channel	2.54*
Equipment: protection channel in trouble	0.5
Fading: two regular channels in simultaneous fades	3.84*
Equipment: protection channel in use by another regular channel	2.0
Total unavailability	9 per cent

* These are average figures. They differ from channel to channel because of differing frequency spacings to the protection channel and to adjacent channels.

the time when protected. Similarly the percentage of the time by which any levels of noise greater than 40 dba are exceeded are reduced to 9 per cent of their unprotected values. The noise distribution to be expected from the ten-repeater switching section of five regular channels is shown in Fig. 3, designated "With Switching."

In a 4,000-mile system made up of such switching sections the noise would exceed 40 dba 0.13 per cent of the time. This figure is for heavy summer night time fading. For evening hours, the percentage would be in the order of 0.024 per cent, and for seasons other than mid-June to mid-September much less. These figures correspond to an equipment failure outage of 0.1 per cent as discussed previously.

Thus the total outage time would be as given in Table II.

CIRCUITS OF THE AUTOMATIC SWITCHING SYSTEM

The circuits of this system can be classified into three main categories:

1. Switching circuits.
2. Evaluating circuits.
3. Control circuits.

The purpose of the switching circuitry is to transfer the TV or telephone signal from an impaired regular radio channel to the protection channel. The evaluation circuits determine when and where the transfer should take place, and when the transfer back to normal should be made. The control circuits are the connecting link between the evaluating and switching circuits and in forming this link, they accept information from the evaluating circuits and cause the switching circuits to operate. At the transmitting end of a switching section the switching is embodied in the transmitting IF switch and at the receiving end it is the receiving IF switch. The evaluating function is located at the receiving end of

TABLE II — PER CENT OF TIME REGULAR CHANNEL UNAVAILABLE
4,000 miles, 5 regular channels, ten repeater switching sections

	No Automatic Protection Switching		With Automatic Protestion Switching	
	Heavy summer night	Heavy summer evening	Heavy summer night	Heavy summer evening
Equipment failure	6.65	6.65	0.1	0.1
Fading	1.46	0.03	0.13	0.02
Total	8.11	6.68	0.23	0.12
Automatic switching improvement factor			35	56

each section and it is performed by a unit called the channel initiator. The way in which these units and the control circuits are connected in a switching section is shown in Fig. 4.

Switching Circuits

At the transmitting and receiving end of a section the switching circuit is embodied in the transmitting and receiving IF switch. There are two units which form the fundamental building blocks in these switches as well as in other switching circuits in TD-2. These are the Bridging Amplifier and the 223-type IF coaxial switch. After describing them we will show how they are put together to form the IF switches.

Bridging Amplifier

The bridging amplifier is a device with a single input and two outputs capable of transmitting a band of 60 to 80 mc. A photograph of this unit is shown in Fig. 5. The path from A to B is a low pass filter with a cut off in excess of 400 mc so that transmission in the 60 to 80 mc band is flat, while the path from A to AB or B to AB passes through a one tube amplifier.

223 Type IF Coaxial Switch

The actual switching is accomplished by means of the 223-type IF coaxial switch. This unit is shown in Fig. 6 and a schematic in Fig. 7. As can be seen from the schematic, this switch has four transfers which operate simultaneously when voltage is applied to the windings. With no voltage across the windings the transmission paths are A to AB and B to

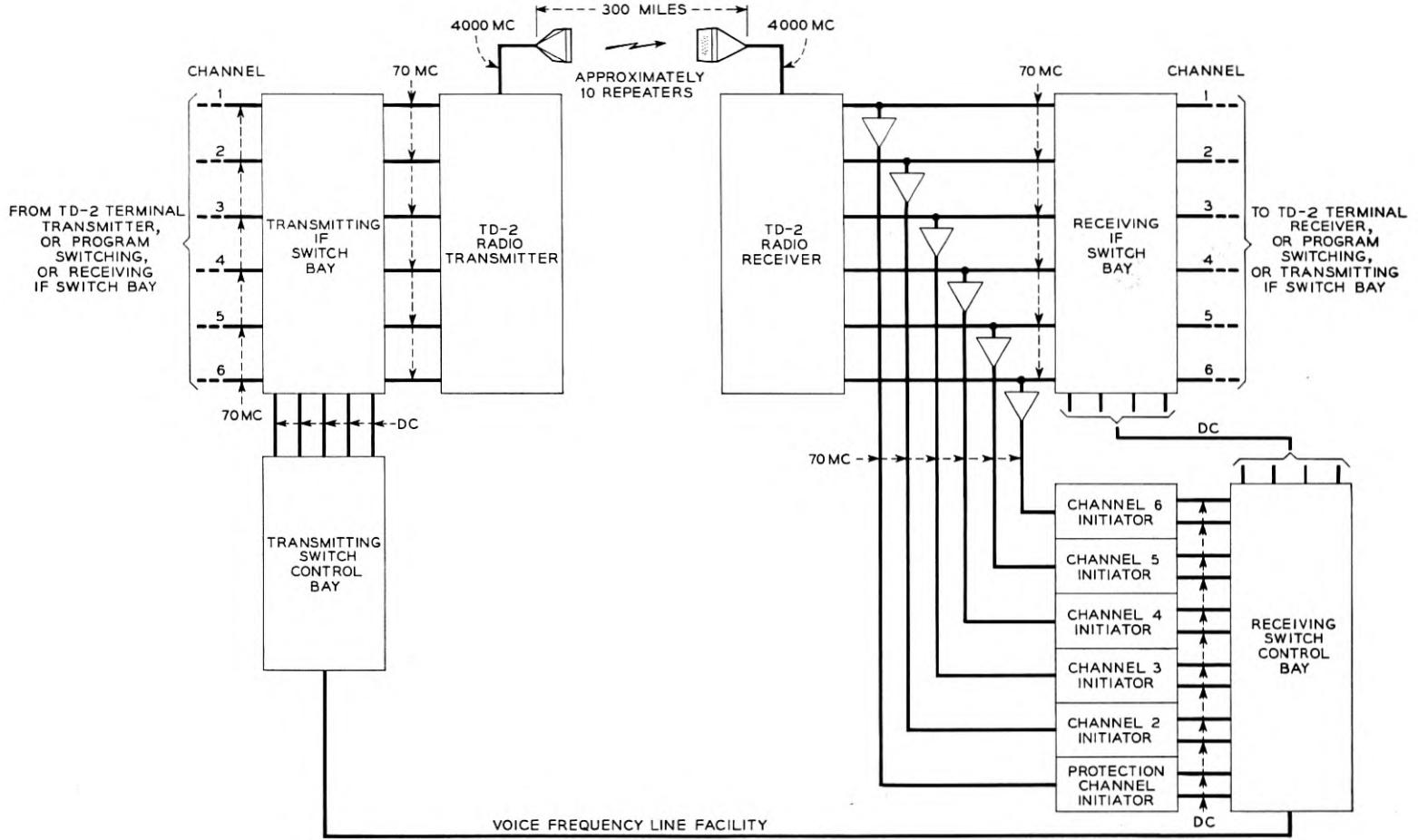


Fig. 4 — Automatic switching.

Fig. 5 — IF bridging amplifier.

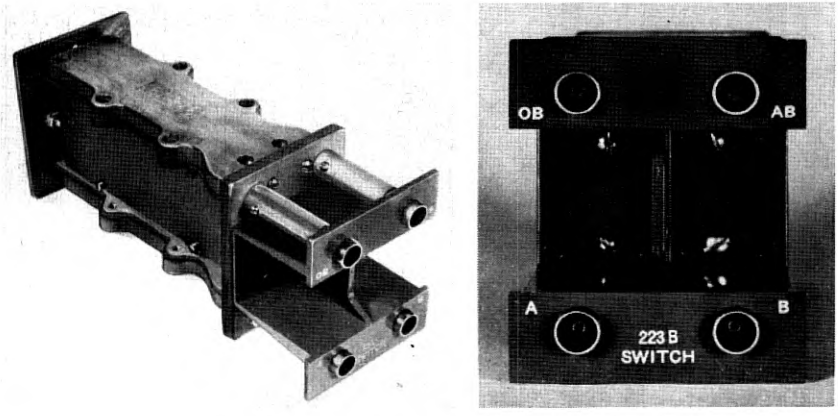


Fig. 6 — 223-type IF coaxial switch.

OB. Where the contacts are operated the transmission paths are A to an internal 75-ohm termination and B to AB. This arrangement of contacts is not the most economical insofar as number of contacts is concerned, however crosstalk between paths must be held to 90 db or greater and this arrangement was found necessary to keep the stray capacity between outputs to a minimum.

The structure of the transmission path in this switch is entirely coaxial and the switches themselves are glass enclosed reeds resting in a pool of mercury. By capillary action the mercury wets the reed and provides a positive contact. The operate time of the switch is 1.5 milliseconds.

Transmitting and Receiving IF Switches

As was mentioned previously, the bridging amplifier and 223 IF coaxial switch are used in combination to perform the job of the IF switching circuit. Fig. 8 presents in schematic form how these units combine to form the transmitting IF switch. Note that the operation of any 223 switch in a regular channel path causes transmission to be put on the regular and protection channel simultaneously. Note also that the active portion of the bridging amplifier is in the normal path. If the tube should fail, then the channel could be restored by an automatic switch. Finally, note the protection channel path which is shown as dotted. An operation of the 223 switch in this path will cause the protection pilot to be removed and the protection channel will be cut through the transmitting switch. This is done when the protection channel is committed to a service other than automatic switching.

A schematic of the receiving switch is shown in Fig. 9. The receiving switch operates under the control of the receiving switch control circuit. From the schematic the following points should be noted. First, the initiators are bridged on their respective channels at all times. A failure of the bridging amplifier will cause an automatic switch to take place

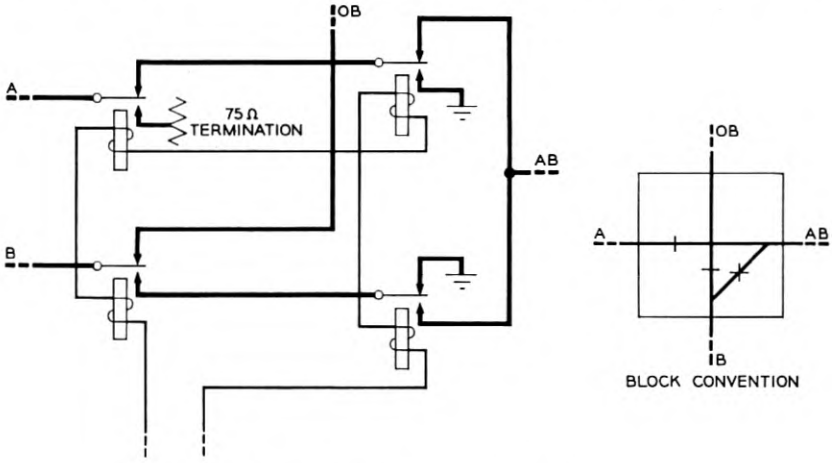


Fig. 7 — 223-type coaxial switch.

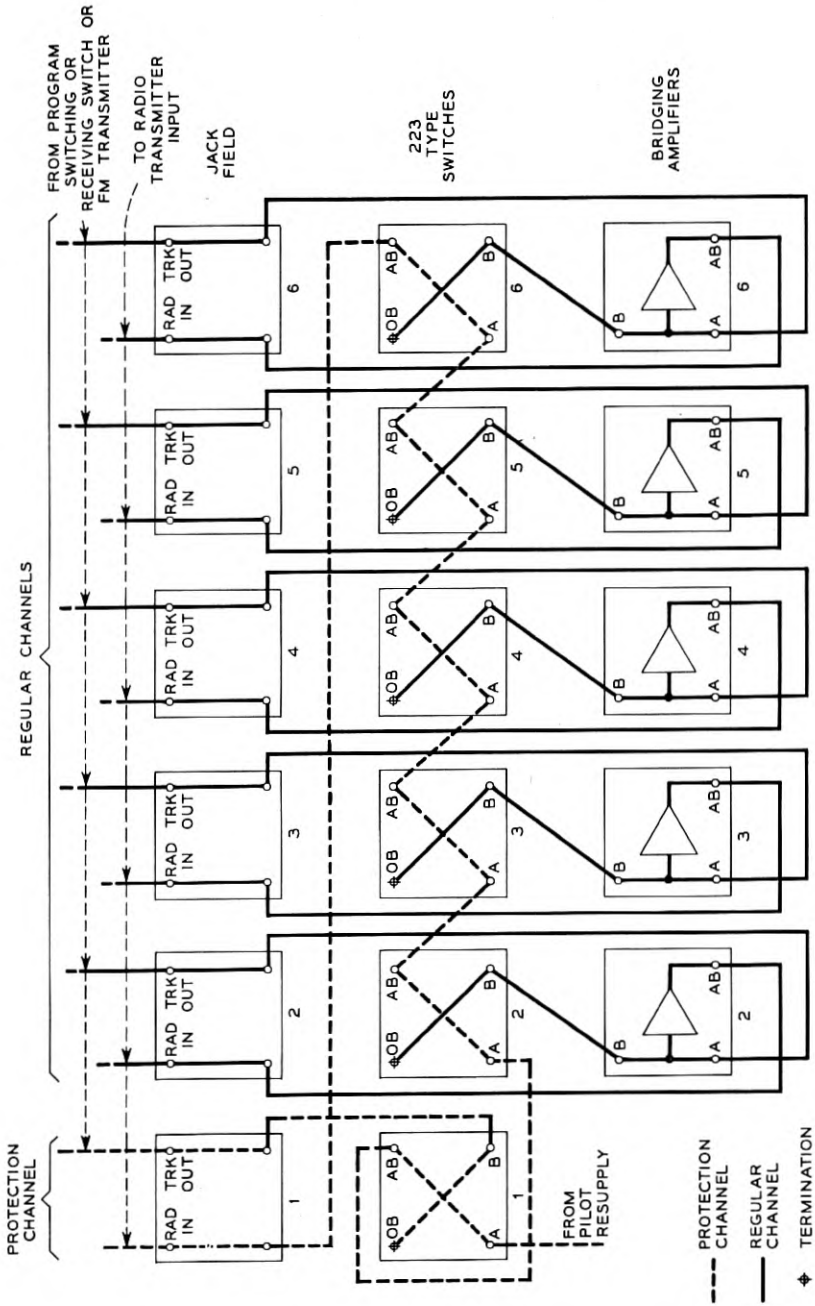


Fig. 8 — TD-2 auto switching t ransmitting switch.

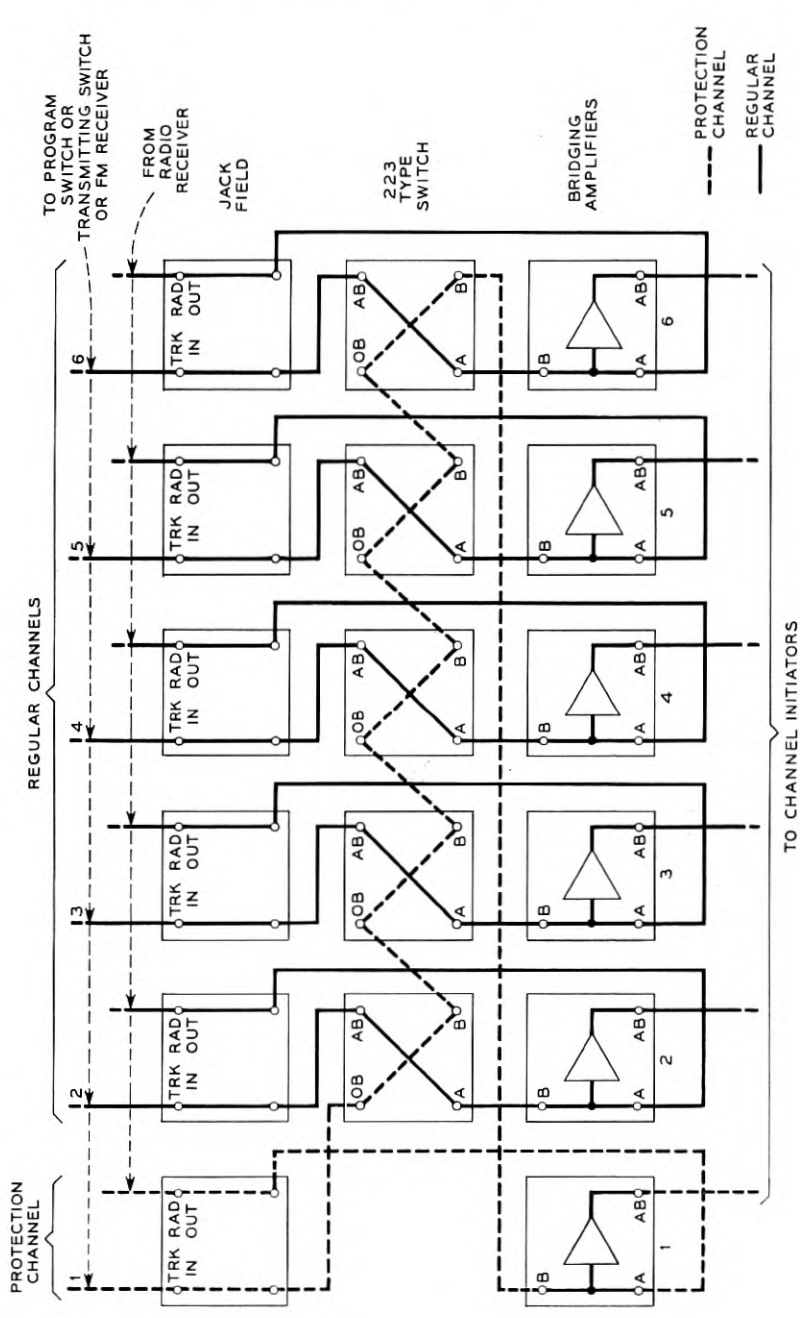


Fig. 9 — TD-2 automatic switching receiving switch.

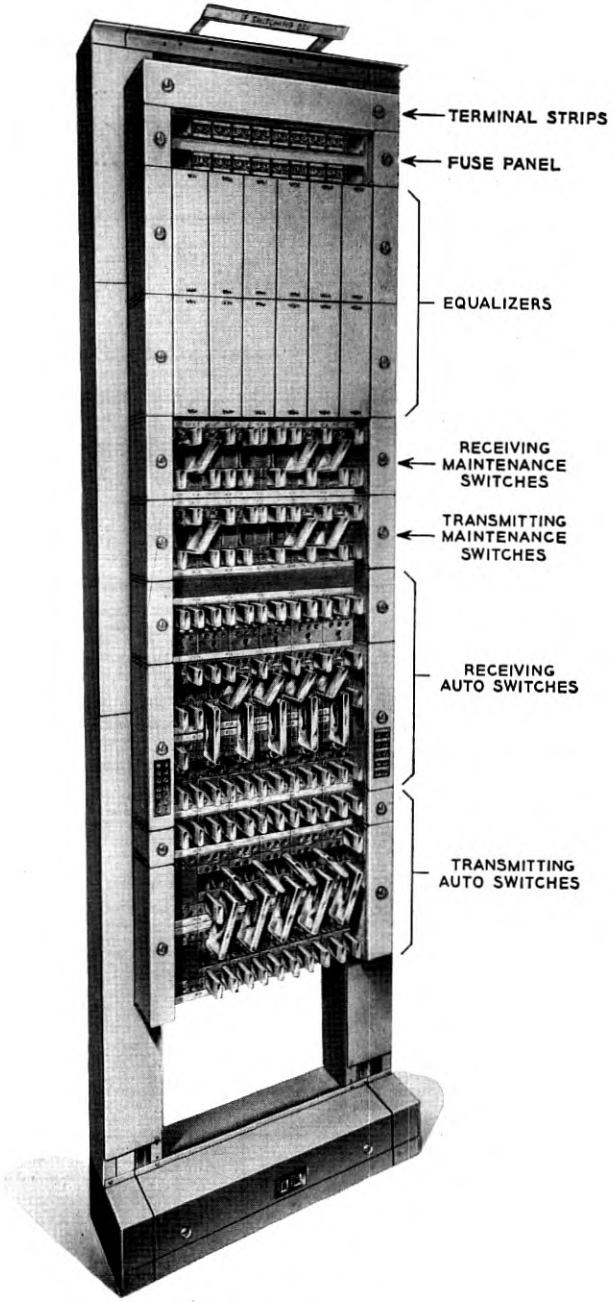


Fig. 10 — IF switching bay.

since the initiator will see no IF carrier. Second, the operation of any 223-type switch causes the incoming regular channel signal to go from the A jack on the 223-type switch to an internal 75 ohm termination. The protection channel is diverted by the operation of the switch and it then becomes the outgoing signal. The IF switching bay is shown in Fig. 10.

Evaluating Circuits

The initiators are the sensing circuits in this system and it is their job to evaluate the condition of a channel. The regular initiators monitor the working channels and by measuring the noise and the carrier level they determine if the channel is in need of protection. The protection initiator monitors the protection channel and by measuring noise and a pilot which is normally transmitted over this channel the initiator can tell if the channel is available for protection service. When a transmitting end switch is made, the protection pilot is removed and is replaced by a TD-2 carrier with message or TV modulation. If this signal is satisfactory the protection initiator then informs the receiving control circuit that a switch should be completed. If the signal is not satisfactory the switch is not completed. We will now describe briefly how the initiators perform their evaluating functions.

Regular and Protection Channel Initiators

A simplified block diagram of the regular initiator is shown in Fig. 11. One input signal to the regular initiator is a 70-mc carrier with frequency modulation in the form of TV or telephone information. Another input is an 8.9-mc regular pilot tone. The pilot, which is locally generated, amplitude modulates the 70-mc carrier so that the output signal of the receiver consists of the TV or message plus an 8.9-mc tone. If the 70-

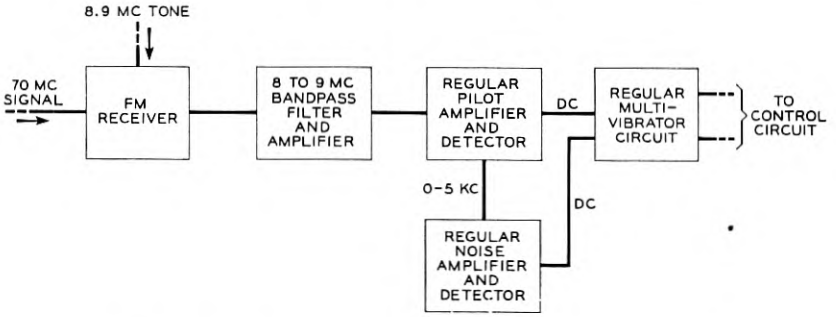


Fig. 11 — Simplified block schematic of regular initiator.

mc carrier is lost due to an equipment failure the 8.9-mc signal at the output of the receiver is also removed. At the output of the receiver is an 8 to 9-mc bandpass filter which removes the TV or telephone information and passes only the pilot tone and any noise around it. The pilot and noise are then amplified in a gain stabilized amplifier and detected. The DC portion of the detected pilot is fed into a bi-stable multivibrator. The ac portion of the detected pilot (which is noise) is amplified and detected. The dc due to the rectified noise is also fed into the multivibrator. Its output is connected to the Receiving control circuit for that particular channel. If the pilot should be removed (due to loss of carrier) or the noise should increase, the multivibrator will change its state and cause the control circuit to operate.

The protection initiator is quite similar to the regular and a simplified block diagram of it is shown in Fig. 12. One of the input signals can be either a 70-mc carrier modulated by an 8.5-mc protection pilot (when no switch has been requested) or a 70-mc signal with telephone or TV modulation (when a switch has been requested). The other input signal is a locally generated 8.9-mc tone, the same as in the regular initiator. This tone, as in the regular initiator, amplitude modulates either of the 70-mc input signals. The output signal of the FM receiver

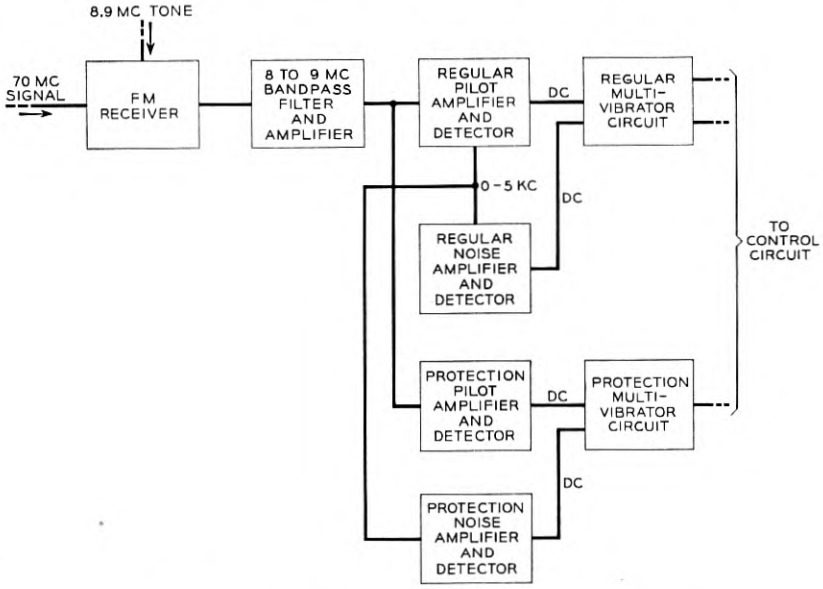


Fig. 12 — Simplified block schematic-protection initiator.

is, therefore, either a pair of tones at 8.9- and 8.5-mc or an 8.9-mc tone with TV or telephone modulation. The block units in the protection initiator are similar in form and function to those in the regular initiator.

At the output of the 8- to 9-mc filter and amplifier the signal is fed simultaneously into regular and protection pilot amplifiers. These amplifiers are tuned to their respective pilot frequencies. The detectors at the output of these amplifiers rectify the pilots and feed the dc voltage to the respective multivibrators. The ac portion (noise) of the regular pilot detector output is fed simultaneously to the regular and protection noise amplifiers. The noise around the regular pilot is used because this tone is constant in level and not affected by transmission vagaries as is the protection pilot. The two noise amplifiers are adjusted so that their gain is 5 db greater than the noise amplifier in the regular initiators. This greater sensitivity in the protection initiator noise amplifier guarantees that a switch will not be completed unless the noise on the protection channel is at least 5 db lower than on the regular channel.

The bi-stable multivibrators in the protection initiator are identical to those in the regular initiators. The output of the protection multivibrator informs the control circuits of the status of the protection channel.

Pilot Supply and Distribution Circuit

The description of the initiators referred to a locally generated 8.9-mc tone. This tone comes from a unit called the pilot supply. Since each initiator in an office requires such a tone, a distribution circuit is provided to divide the high level output of the pilot supply into a maximum of twenty equal low level tones.

The pilot supply consists of two identical units each having a crystal controlled oscillator, amplifier and limiter, and a single transfer circuit. The limiter serves two functions — it maintains the output at a constant level and the limiter current is used to operate the transfer circuit. Each of the units operates continuously — one serving as a standby. Their outputs feed into a 223 type coaxial switch. If one of the units fails, the transfer circuit sees a loss of limiter current from that unit and transfers the working unit to the output by operating the 223-type switch.

Protection Pilot Resupply Circuit

In our description of the system we have made frequent mention of the protection pilot. In our earlier remarks we also described how the TD-2 system was divided into switching sections of approximately 10

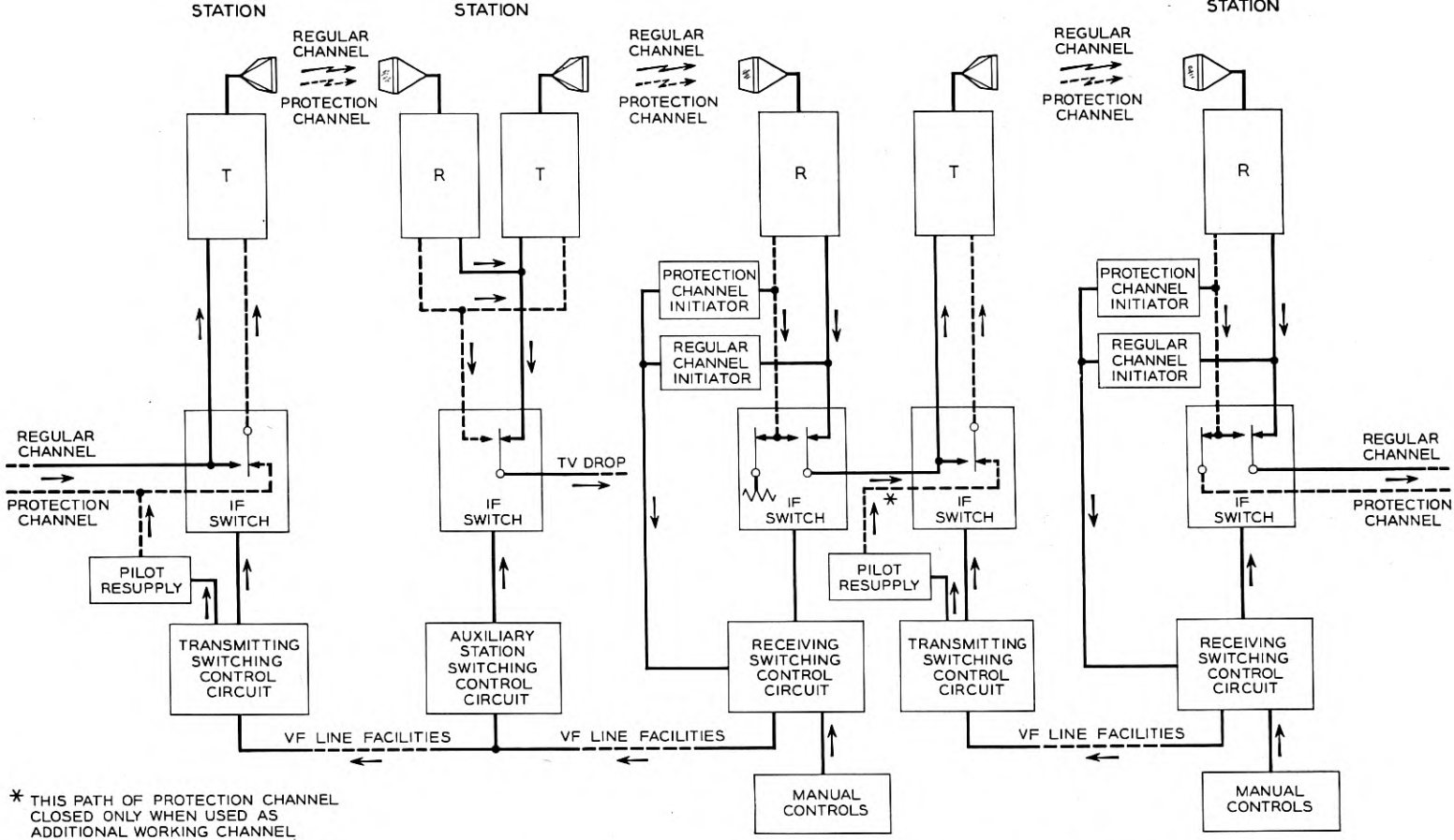


Fig. 13 — Control circuits required for one direction of transmission for two sections of a TD-2 system.

repeaters. At the transmitting end of each section the protection pilot is inserted. One method of providing this pilot on the channel would be to have an 8.5-mc tone modulate a terminal transmitter. This becomes uneconomical since a terminal transmitter would have to be provided for each switching section. A signal which is equivalent to a carrier, frequency modulated by a single tone, can be obtained by adding two tones of the correct level and frequency and then passing the resultant through a limiter. A derivation in the appendix shows the requirements on level and frequency for this equivalence.

CONTROL CIRCUITS

The control circuits accept the signals from the channel initiators and cause the IF switches to operate and replace a regular channel by the protection channel. Fig. 13 shows in block form the arrangements of the control circuits required for one direction of transmission for two sections of a TD-2 system. Following is a brief discussion of the operation of the control circuits for the various transmission conditions of a system.

Regular Channel Failure

When all the regular channels of a section are working properly, a 700-cycle guard tone is transmitted over the voice frequency path from the receiving switching control circuit to the transmitting switching control circuit which holds the IF switches in a released condition. The purpose of the 700-cycle guard tone is to prevent the operation of the switches on extraneous signals, such as noise on the VF line, and to provide an indication that the voice frequency facility has not been impaired. When transmission on any one of the regular channels is impaired, the initiator associated with that channel sends a signal to the receiving switching control circuit. If the protection channel is available, the receiving circuit removes the 700-cycle guard tone, applies another tone depending on the channel in trouble, and enables the associated receive IF switch but does not operate it. Each channel has a discrete identification tone and the frequencies used for the five regular channels are 900 cycles, 1,100 cycles, 1,300 cycles, 1,500 cycles and 1,700 cycles. On the removal of the 700-cycle guard tone and the reception of the channel identification tone, the transmitting switching control circuit bridges the protection channel across the impaired regular channel and removes the 8.5-megacycle pilot frequency from the protection channel. The regular channel IF carrier, which is now applied

to the protection channel, causes the protection channel initiator at the receive end of the section to operate the previously enabled IF receive switch and transfer the outgoing regular channel to the protection channel, thereby completing the replacement of the impaired channel by the protection channel.

The IF switches at both ends of the section remain operated until the transmission of the replaced regular channel returns to normal at which time the initiator at the receive end of the circuit sends a signal to the receiving switching control circuit to release the receive IF switch and reconnect the regular channel through to the succeeding section. At the same time, the channel identification tone is removed from the voice frequency path and the 700-cycle guard tone reapplied. This results in the restoration to normal of the transmitting switching control circuit and the subsequent release of the transmit IF switch. When this switch releases, the bridge between the regular channel and the protection channel is opened and the pilot resupply circuit reapplies the 8.5-mc tone to the protection channel, thereby completing the restoration of all circuits to normal. However, during the period of time between the release of the receive IF switch and the reception of the protection channel 8.5-mc pilot tone, the receiving switching control circuit locks-out requests for the protection channel by all other regular channels. This feature is provided to assure that the receiving control circuit will not attempt to complete a new switch until all circuits at the transmit end of the section have returned to normal.

When the transmission of two or more channels in the same section is impaired at approximately the same time, the initiators of these channels will cause the receiving switching control circuit to apply tone toward the transmit end for each of the channels. No action will be taken at the "transmit" end at this time as the transmitting control circuit is designed to ignore the request for a switch when more than one channel identification tone is received. At the receive end, the control circuit locks out all but the lowest numbered channel. When the lock-out is completed, the tone of the lowest numbered channel only will be present on the voice frequency line and the transmitting control circuit will then operate the IF switch for that channel. If due to a trouble condition, the switch cannot be completed on the above mentioned lowest numbered channel within 50 milliseconds, that channel will be locked out by the receiving control circuit for a period of 10 seconds and the control circuits will try to complete a switch for the next lowest numbered channel requesting a switch. If a switch is completed all other channels will then be locked out until the protection channel

is again available, but if a switch cannot be completed for any of the channels the receiving switching control circuit will try again, at the end of the 10-second period mentioned above, to complete the switch for the lowest numbered channel then requesting a switch.

Voice Frequency Path or Protection Channel in Trouble

The above general discussion of the operation of the control circuits is based on the assumption that the voice frequency path between the control circuits is good and that the protection channel transmission is satisfactory. It is obvious that these conditions may not exist at all times and the control circuits have been designed accordingly. If the transmission of the voice frequency path fails after a switch has been completed, the switch will remain operated even though the transmitting switching control circuit no longer receives the channel identification tone. This is accomplished by a feature of the control circuit which recognizes the absence of both the guard tone and the channel tone, and which causes the switch to remain in the condition existing just prior to the loss of transmission on the voice frequency path.

If the transmission of the protection channel is impaired, the absence of the 8.5-mc pilot tone will cause the receiving switching control circuit to lock out requests from all regular channels and thereby will prevent the start of any switch.

Switch Request by Channel in Preceding Section

When a regular channel is impaired, a switching operation will be started on this channel in the succeeding switching sections provided the protection channel is available or not in trouble. When this occurs, the IF switches at the transmit ends will operate in these sections, but the IF switches at the receive ends will not operate. This is because the protection channel initiator will not see a good regular channel signal since the fault is in a preceding section. When the switch is completed in the section in trouble, the transmission of all the succeeding sections will return to normal with the result that all the control circuits of these sections will return to the condition dictated by the transmission within the respective sections. If, however, a switch cannot be completed in the section in trouble, the transmit switches in all succeeding sections as well as the section in trouble will remain operated for a period of approximately 50 milliseconds. At the end of this time the receiving switching control circuit of each section will lock out the channel in trouble for 10 seconds. The lock-out will remove the channel identifica-

tion tone from the voice frequency paths and will release the transmit IF switches. At the end of the 10-second lock-out period, the receiving switching control circuit will again attempt to complete the switch.

Auxiliary Switching Stations

There are network broadcast conditions which require that an auxiliary station provide transmission to locations not on the path followed by the "back-bone" or main line. An auxiliary station is equipped with a transmitting switching control circuit essentially the same as that provided at a transmit station except that a 2,100-cycle identification tone receiving circuit is furnished in addition to the identification tone receiving circuits for the regular channels. However, at the auxiliary station, the operation of the control circuit differs in that although it responds to the removal of the guard tone and the presence of the channel identification tone, it does not complete the switch. When the switch for the back-bone circuit is completed at the receive end of the circuit, the receiving switching control circuit transmits a 2,100-cycle tone over the voice frequency path to the auxiliary station. The control circuit at the auxiliary station then recognizes this tone and completes the previously enabled switch. If, after the switch has been completed the 2,100-cycle tone is removed from the voice frequency path, the switch will be maintained until the 700-cycle guard tone is reapplied to return the circuit to normal.

The presence of the 2,100-cycle tone on the voice frequency path does not influence the control circuit at the transmit end as it is outside the frequency limits of the filters used in the transmitting switch-control circuit.

Manual Controls

Although the switching system has been designed primarily for automatic operation there are occasions when manual switching is required. Typical conditions under which manual switching is used, and the operation of the control circuits for these conditions are as follows:

Manual Switch

By the simultaneous operation of a manual switch key and a master key, located at the IF switching bay or at a remote point, the receiving switching control circuit will start a switch just as if an initiator had requested a switch, except that the 50 millisecond timer will be disabled.

All regular channels and all other positions from which a manual switch may be started will be locked out. If the protection channel is not in use, the manual switch will be completed at both ends of the circuit in the same manner as if an automatic switch had been requested. A visual indication that a manual switch has been started is given by the lighting of a lamp at each location from which a manual switch may be initiated.

If the protection channel is in use for another regular channel at the time the manual switch is started, no action will be taken by the receiving switching control circuit as all manual positions will be locked-out.

Regular Channel Lockout

There are times when locking-out an impaired regular channel will give operating advantages by releasing the protection channel for use with the remaining regular channels. The lock out feature is accomplished by the simultaneous operation of a master key and the channel lock-out key.

Forced Manual Switch

Whenever a switch is completed, either automatically or manually, a lamp will light at all associated manual control positions. If a manual switch has been started and is not completed because of unsatisfactory transmission over the protection channel, the lamp will not light. However, as previously mentioned, the 50-millisecond timer is disabled when a manual switch is attempted and therefore the protection channel will remain bridged across the impaired regular channel at the transmit end of the circuit. This will permit transmission tests to be made on the protection channel and if these tests indicate that the protection channel transmission is suitable for service, the switch can be forced or completed manually by the simultaneous operation of the master key and the forced switch key. A channel lamp will light to indicate that the forced switch has been ordered.

Protection Channel Lockout

Traffic may demand that the protection channel be locked out of the automatic switching system and be made available as an additional working channel. This can be accomplished by the simultaneous operation of the master key and the lockout key. At this time, if the protection channel is available, the receiving switching control circuit locks out all the regular channels and transmits to the transmitting switching

control circuit the identification tone of the lowest numbered regular channel in addition to the 700-cycle guard tone. The transmitting switching control circuit recognizes the presence of both tones and removes the 8.5-mc pilot tone from the protection channel. A visual indication is given at the transmit end to show that the Pilot Resupply relay is operated and that the pilot tone is not applied to the protection channel. At the receive end of the circuit, a lamp will light to indicate that the lock out has been completed.

If the protection channel is not available when the proper keys are operated, a lamp will light at the receive end to indicate that the lock-out has not been completed. When the protection channel becomes available, the circuits will function as described above.

Alarms

The importance of each of the regular channels and of the protection channel makes it apparent that alarm features must be provided to indicate the failure of any channel or of the automatic switching system.

An automatic switch alarm circuit in conjunction with the receiving switching control circuit provides the following alarms:

1. A major alarm is operated when a regular channel fails and is not replaced by the protection channel within a few seconds.
2. Minor audible and visual alarms are operated to indicate failures on the regular and protection channels when the failure lasts for more than 15 to 45 seconds. The visual alarm indicates the number and direction of the channels in trouble.
3. A major alarm is also operated within a few seconds if the protection channel fails after it has been switched for an impaired regular channel or as an additional channel.
4. A major alarm is operated if a pilot re-supply circuit which provides the 8.5-mc tone for the protection channel fails, or if both generators in the pilot supply fail.
5. Minor audible and visual alarms are operated to indicate failure of a voice-frequency path. Two alarm lamps are lit for a failure. One lamp indicates that the voice-frequency path is in trouble and the other lamp indicates the direction of transmission.

EQUIPMENT ARRANGEMENTS

The automatic switching equipment is grouped together on a terminal basis. A terminal group for a switching section includes the receiving switching equipment for all incoming channels from a specific

point and also the transmitting switching equipment for all outgoing channels in the opposite direction to that point. Two such terminal groups, connected on a back-to-back basis, are required at an intermediate switching stations having two directions of transmission. When program switching facilities are required they are connected between the automatic switching groups.

The equipment used in both automatic and program switching is mounted in 9-foot high, 19-inch wide duct type bays. These bays are the same type as used on present TD-2 radio equipment. In each terminal group for automatic switching there are three basic bays. These are: The IF switching bay, the initiator bay, and the switching control bay.

The IF switching bay contains the receiving and transmitting IF switches for a particular direction. The capacity of this bay is six channels incoming and six channels outgoing (5 regular and 1 protection).

The initiator bay provides for a maximum of five regular initiators and one protection initiator. Provision is also made for an additional protection type initiator which may be used as a spare on regular or protection channels. The bay is arranged for double side maintenance with controls, meters, tube and jack appearances on the front or operation side. This bay is shown in Fig. 14.

There are two standard bay arrangements provided for the switching control equipment. One arrangement is used at the receiving and transmitting ends of a section and the other at an auxiliary station. The control equipment not mounted in these bays are the pilot resupply circuit which mounts on a miscellaneous basis, the alarm lamps which mount on a miscellaneous basis, and the manual switch controls which are mounted on the IF switch bay.

Switching Control Bays

The switching control bay, Fig. 15, accommodates the transmitting switching control equipment serving two to six radio channels to a particular point and the receiving switching control equipment serving two to six radio channels from the same point. The equipment mounted in the bay is as follows:

1. One *receiving switching control group unit* which provides basic control equipment for the receiving end of a switching section. This unit sends out the guard tone and also receives information from the protection channel initiator. A second oscillator mounts on this unit to supply the 2,100-cycle signal for auxiliary station drops.

2. A *receiving switching control channel unit* for each of the regular

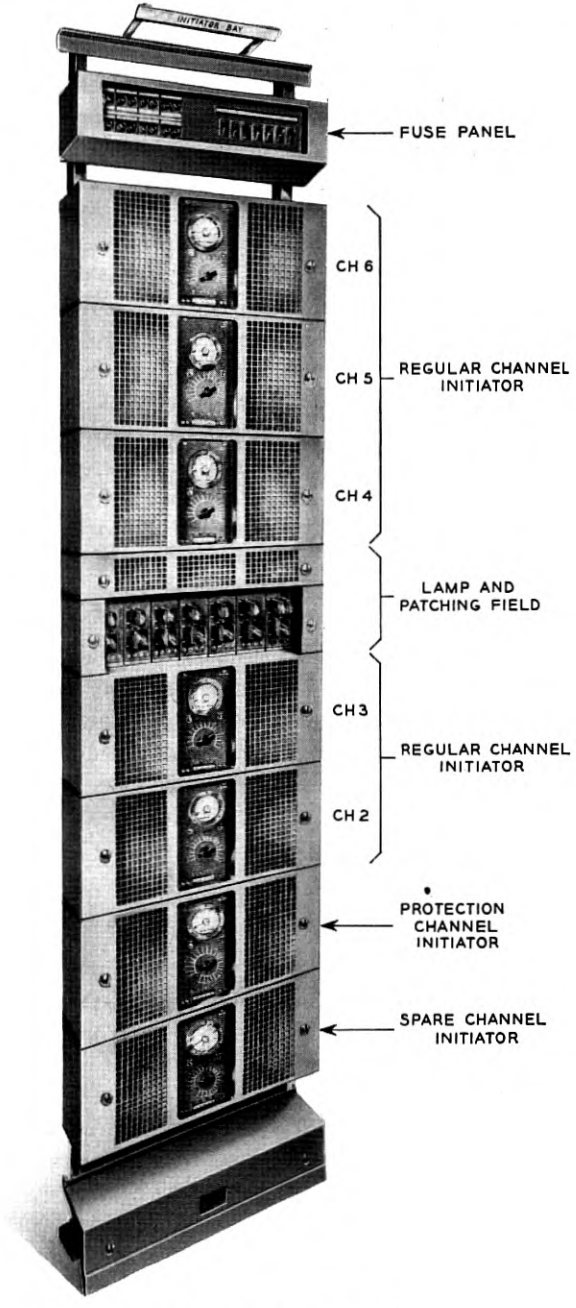


Fig. 14 — Initiator bay.

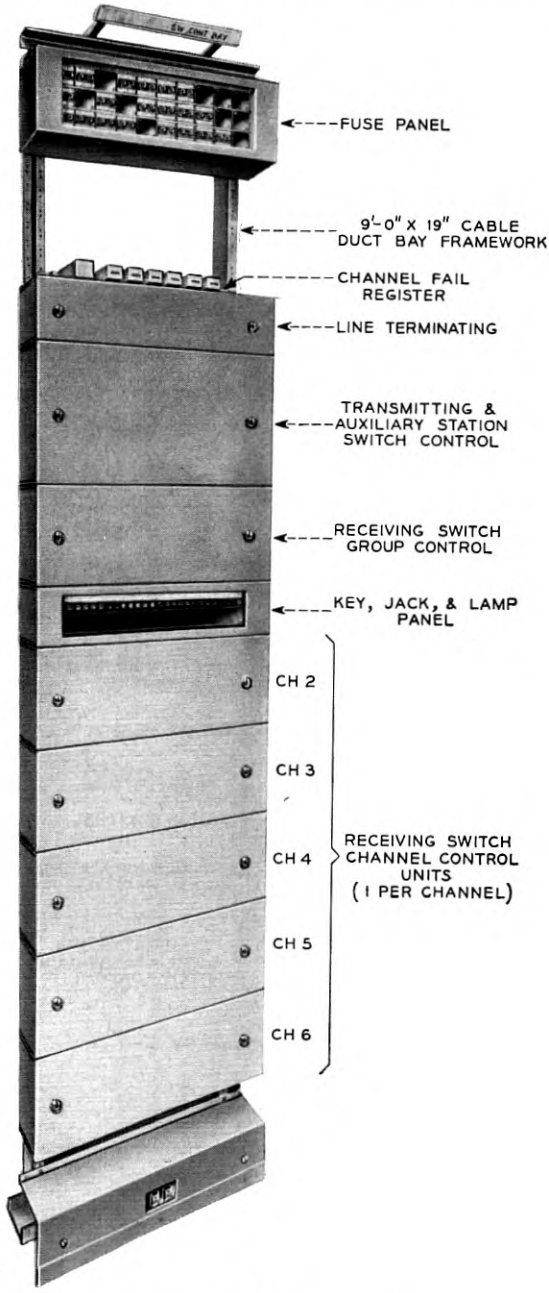


Fig. 15 — Switching control bay.

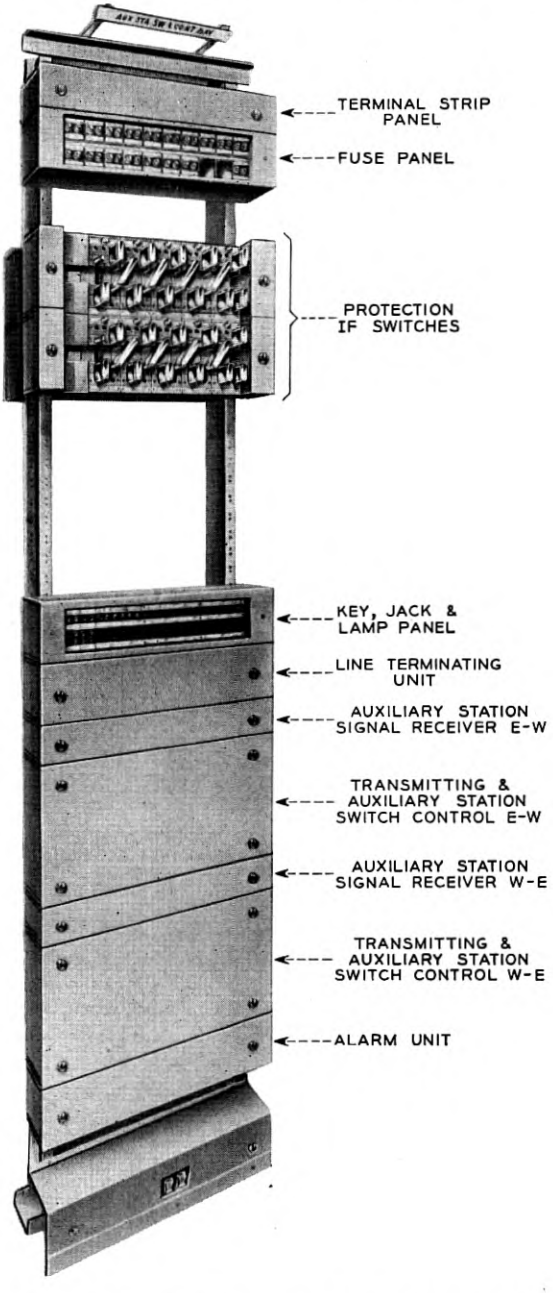


Fig. 16 — Auxiliary station switching and control bay.

channels at the receiving end of a switching section. This unit receives information from the regular channel initiators and sends out the identifying channel tones.

3. One *transmitting and auxiliary station switching control unit* which provides basic control equipment for the regular channels at the transmitting end of a switching section. This unit receives and amplifies the identification and guard tones.

4. One *channel detector unit* which provides the filtering and detectors for the channel identification signals.

5. One *line terminating unit* required for the line facilities over which the VF signals are sent.

6. *Key, jack and lamp panel.* This panel provides a mounting for various indicating lamps, line terminating test jacks, order wire jacks and a 24-volt test battery jack for checking relay equipment.

7. *Channel failure registers.* This unit provides arrangements to count the number of times that the protection channel transmission falls below the allowable minimum and also to count the number of times the protection channel is substituted for each of the regular channels.

Auxiliary Station Switching and Control Bay

The auxiliary station switching and control bay, Fig. 16, provides the equipment required for customer drop protection at auxiliary stations. The equipment mounted in the bay is as follows:

1. The *IF switches* which transfer the customer drop from the regular channel to the protection channel.

2. A *transmitting and auxiliary station switching control unit and channel detector units* discussed above in Items (3) and (4) of the switching control bay.

3. An *auxiliary station signal receiver unit* for each direction to receive the 2,100-cycle signal sent from the receiving station.

4. An *alarm unit* with arrangements for either remote or local operation.

5. A *key, jack and lamp panel* to mount the various alarm lamps, keys, test jack, order wire jacks and a 24-volt test battery jack for relay testing.

6. A *line terminating unit* to terminate the VF line between the backbone circuit and the auxiliary stations over which the signals to control the switch are sent.

PROGRAM SWITCHING

We have so far described the operation of automatic switching and the circuits used to provide automatic protection. In a system such as the

TD-2 where a large number of broadcasters are provided with TV service, a manual switching arrangement is necessary and program switching performs this function.

In transmitting TV a great deal of flexibility is needed to meet the needs of service. A few examples of when such flexibility is desirable are when the protection channel is used for service other than automatic switching or when a particular program must be transmitted over a number of channels simultaneously.

Program switching is done outside the framework of automatic switching. Fig. 17 shows in block diagram form how program switching ties in with TD-2 and automatic switching.

To meet the great variety of situations which arise program switching has to be engineered to meet the needs of each office individually. In order to have the necessary versatility this system has been designed around three basic circuits shown in Fig. 18. These are:

a. Through channel group.
b. Bridging amplifier group.
c. Switch group.

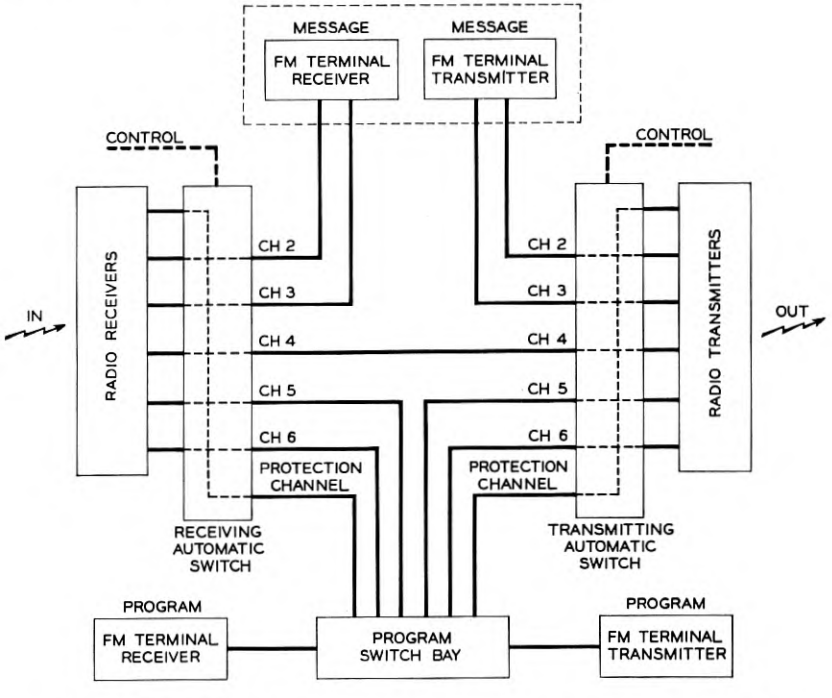


Fig. 17 — TD-2 radio automatic and IF program switching systems.

The thru channel group shown schematically in Fig. 18 provides a passive circuit for through transmission; an outlet for connection to branch circuits and a switch which can break the through path and feed the outoing channel with a local signal.

The bridging amplifier group contains one or more bridging amplifiers. The number of outputs available is one greater than the number of amplifiers in the group.

The switch group operates to connect any one of a number of inputs to any one output or it may connect any single input to any one of a number of outputs.

With these three units the desired flexibility of program switching can be achieved. An example of how they may be used in combination is shown in Fig. 19 which shows a simple example of program switching. In actual offices where there may be four directions of transmission as well

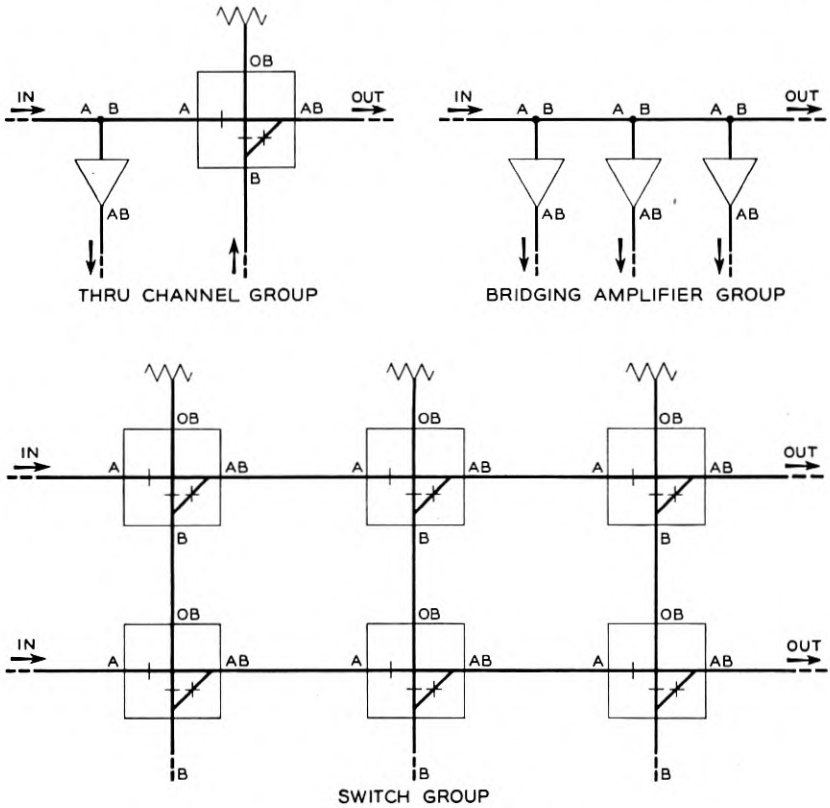


Fig. 18 — Basic circuits for program switching.

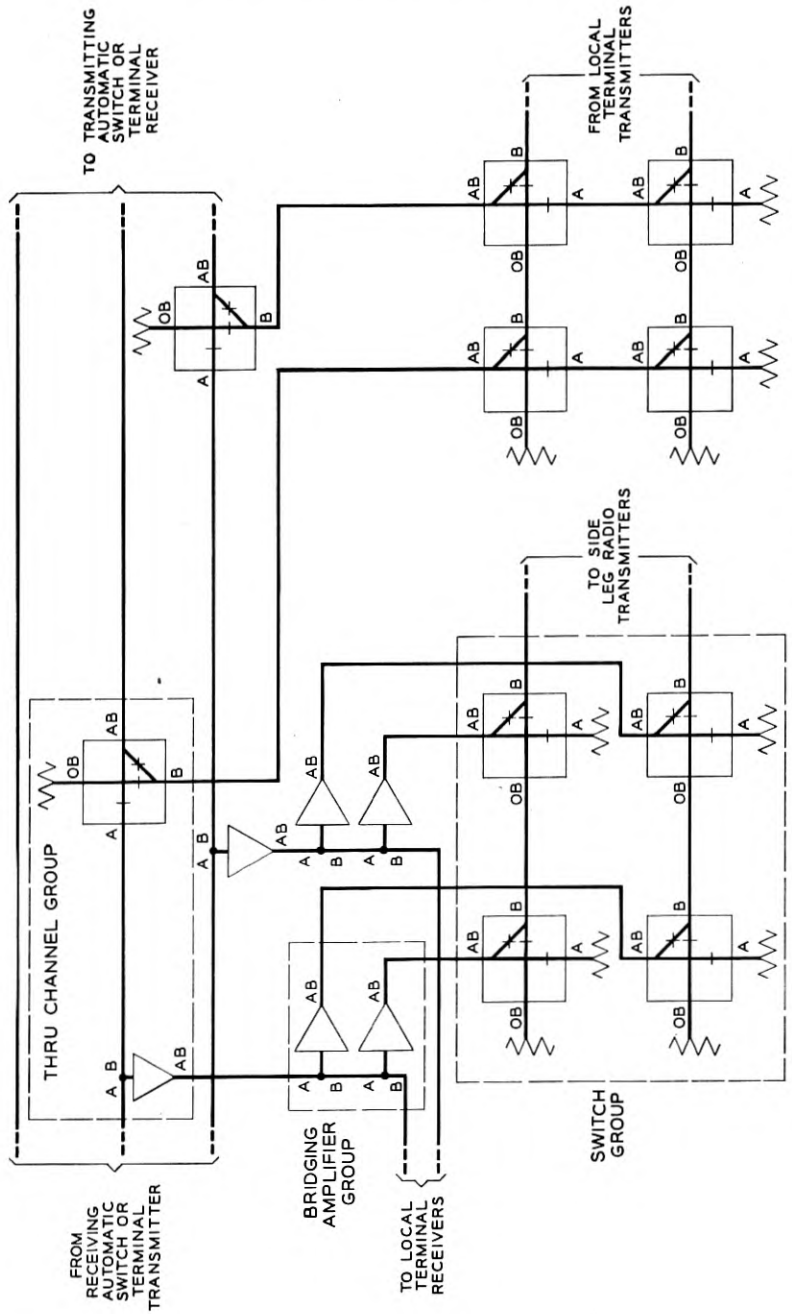


Fig. 19 — Example of program switching.

as a number of drops and pick-ups the circuits become much more intricate.

The operation of these switches can be performed locally by push buttons or from remote locations.

CONCLUSION

The TD-2 automatic protection switching system has been in service between Denver and Sacramento since mid-June 1954. Its performance has been followed closely by special evaluation test equipment installed at each switching station to record on message registers:

1. The number of channel initiator requests.
2. The number of seconds of channel initiator requests.
3. The number of switches made in each channel.
4. The number of seconds that switches are made in each channel.
5. The number of trouble conditions in the protection channel and the number of seconds that trouble conditions persisted.

In addition strip-chart recorders record each operation of each channel initiator and each switch, so that specific events can be analyzed in detail. This equipment does not evaluate the ability of the channel initiators to detect trouble conditions, but observations of message and television circuit performance have indicated that they are performing in the manner intended.

This system comprises 33 repeater sections arranged in three switching sections of 15, 7 and 11 repeater sections and has six westbound and five eastbound radio channels. It traverses difficult terrain, with many of the stations on mountain top locations, so that the outage time per equipment trouble is higher than in other parts of the country. There is, however, comparatively little fading in this area, probably because of the low moisture content of the air.

In the month of August, 1954, a typical month, switches were requested for 117,357 seconds, and 94.8 per cent of this time resulted in a switch. Of the 6,116 seconds of time requested but not switched, 1,974 seconds were not switched because the protection channel was in use by another channel, and the remainder were not switched because of other reasons including protection channel trouble.

These tests, together with service observations, lead to the conclusion that the Automatic Protection Switching System is effecting a substantial and satisfactory improvement in the reliability of TD-2 radio systems. Accordingly, the switching system is being installed on most of the important back-bone TD-2 routes.

Appendix I

The probability of interruption of a regular radio channel because of equipment failure is derived below. Regular channels are those used for telephone or television business; the protection channel is reserved for use instead of any regular channel which has failed or faded. Fading is considered separately from equipment failures.

We will consider a long radio system divided into switching sections of n repeater sections each. In practice the number of repeater sections per switching section will vary according to the geography of the route, but the argument below is easily extended to this situation. We will consider that there are N regular channels and one protection channel, and thus the total number of radio channels is $N + 1$.

1. Probability that a one-way repeater will be in a failed condition (i.e., the time that it is out of service in a year divided by the total time in a year) $= p$.

2. Probability that a specific one-way channel will be in a failed condition in a switching section of n repeaters very closely $= np$.*

3. Probability that a specific one-way channel and its protection channel both will be in a failed condition within a switching section of n repeaters (i.e., the time that both are out of service simultaneously divided by total time) very closely $= (np)^2$.

4. Probability that any specific one-way regular channel and some one of the other regular channels will be in a failed condition within a switching section of n repeaters very closely $= (N - 1)(np)^2$ where N is the number of regular channels.

5. But one of the two failures predicted in (4) will be made good by the protection channel, and therefore the probability that a specific one-way channel will be interrupted because of any one of the other regular channels has failed very closely $= \frac{1}{2}(N - 1)(np)^2$.

6. Then from (3) and (5) the probability that any specific one-way channel will be interrupted because of failure of any two channels very closely

$$= (np)^2 + \frac{(N-1)(np)^2}{2} = \frac{(N+1)(np)^2}{2}.$$

* More rigorously this probability

$$= 1 - (1-p)^n = 1 - \left(1 - np + \frac{n(n-1)}{2}p^2 - \cdots\right)$$

$$= np - \frac{n(n-1)}{2}p^2 + \cdots$$

where the terms subsequent to np represent the probability that the channel will be interrupted because more than one repeater has failed in the same channel in the same switching section. Since p is small, these terms can be neglected. The same argument applies to the discussion as devoted by the words "very closely."

7. Since p is small, the probability of three or more failures is very small and can be ignored.

8. Then the improvement in reliability of a one-way channel effected by automatic switching is the ratio of the probability of interruption without automatic switching to the probability of interruption with automatic switching and very closely

$$= \frac{2np}{(N+1)(np)^2} = \frac{2}{(N+1)np}$$

9. If a system is made up of R repeaters divided into switching sections of n repeaters, the probability that any specific one-way channel will be interrupted somewhere in the system because of failure of any two channels in any switching section very closely

$$= \frac{R}{n}\frac{(N+1)}{2}(np)^2 = R\frac{(N+1)}{2}np^2$$

(This neglects a small probability of simultaneous failure in more than one switching section of the system.)

10. The improvement effected by automatic switching in the reliability of a one-way channel then very closely

$$= \frac{2Rp}{R(N+1)np^2} = \frac{2}{(N+1)np}$$

which is the same as (8) above.

11. Telephone circuits comprise two directions of transmission, and interruption to either direction of transmission disrupts the circuit, so that for telephone the probability of interruption must be multiplied by two and the improvement divided by two.

Appendix II

The output signal of a TD-2 terminal transmitter when modulated by protection pilot is:

$$\begin{aligned} E_{\text{out}} = {} & J_0(x)\sin\omega_0 t + J_1(x)[\sin(\omega_0+\omega_p)t - \sin(\omega_0-\omega_p t)] \\ & + J_2(x)[\sin(\omega_0+2\omega_p)t + \sin(\omega_0-2\omega_p)t] \\ & + \cdots \end{aligned}$$

where: $J_n(x)$ are the Bessel Functions of the first kind

ω_0 = the carrier angular frequency $(2\pi \times 70 \times 10^6)$

ω_p = the protection pilot angular frequency $(2\pi \times 8.5 \times 10^6)$

x = Index of modulation $= \frac{\Delta\omega}{\omega}$

Since x is small (approximately .1) only the carrier and first order sidebands are necessary and $J_0(x)$ and $J_1(x)$ can be replaced by 1 and $x/2$ respectively. Therefore:

$$E_{out} = \sin \omega_0 t + \frac{x}{2}[\sin (\omega_0 + \omega)t - \sin (\omega_0 - \omega)t]$$

$$E_{out} = \sin \omega_0 t + x \cos \omega_0 t \sin \omega t$$

To a first approximation a signal of this form can be generated by adding two tones and using only the phase of the resultant. A requirement that one tone must be much larger than the other must also be met. As our tones we will take an output of the form:

$$\begin{aligned} E_{out} &= a_1 \sin \omega_0 t + a_2 \sin (\omega_0 - \omega_p)t \\ &= A \sin (\omega_0 t - \varphi) \end{aligned}$$

where $$A = \sqrt{a_1^2 + a_2^2 + 2a_1 a_2 \cos \omega_p t}$$

and $$\varphi = \tan^{-1}\left(\frac{a_2 \sin \omega_p t}{a_1 + a_2 \cos \omega_p t}\right)$$

We now impose the restriction that $a_1 \gg a_2$. Therefore we can approximate φ with:

$$\varphi \approx \frac{a_2}{a_1} \sin \omega_p t$$

The sum of the two tones is then:

$$E_{out} = A \sin\left(\omega_0 t + \frac{a_2}{a_1} \sin \omega_p t\right)$$

$$E_{out} \approx A\left[\sin \omega_0 t \cos\left(\frac{a_2}{a_1} \sin \omega_p t\right) + \cos \omega_0 t \sin\left(\frac{a_2}{a_1} \sin \omega_p t\right)\right]$$

Using again the fact that $a_2/a_1 \ll 1$

$$E_{out} \approx A\left[\sin \omega_0 t + \frac{a_2}{a_1} \cos \omega_0 t \sin \omega_p t\right]$$

When this wave is passed through a limiter, the amplitude variation is removed. We can thus generate an FM signal by adding a pair of tones — a high level 70-mc and a low level 61.5-mc tone. The difference in level of these tones determines the equivalent index of modulation.

REFERENCES

1. A. A. Roetken, K. D. Smith and R. W. Friis, The TD-2 Microwave Radio Relay System, B. S. T. J., **30,** Oct. 1951.
2. A. B. Crawford and W. C. Jakes, Jr., Selective Fading of Microwaves, B. S. T. J., **31,** Jan., 1952.
3. R. L. Kaylor, A Statistical Study of Selective Fading of Super-High Frequency Radio Signals. B. S. T. J., **32,** Sept., 1953.

Transmission Of Digital Information over Telephone Circuits

By A. W. HORTON, JR., and H. E. VAUGHAN

(Manuscript received January 14, 1955)

The problems of transmission of digital data over commercial telephone lines are studied with particular attention to transmitting telephone numbers over a wide variety of circuits. The electrical characteristics of these lines and the noise present on them are major factors which influence the design of a reliable high-speed data transmission system. A system which takes account of these factors has been designed and tested. The results of tests indicate that it operates reliably on almost all message circuits at a rate of about 650 bits per second with signals occupying the band from 700 to 1,700 cycles.

INTRODUCTION

Transmission of digital information between central offices in the same or different cities is a basic process necessary to establishing a connection between subscribers. The existing methods of employing conventional dial pulses at the rate of about 10 pulses per second, or multifrequency pulsing at about 8 digits per second are adequate when used between existing central offices. When consideration is given to the application of electronic techniques to central office switching, it becomes apparent that higher signaling speeds are desirable if not necessary.

The work reported here was carried out several years ago with the objective of selecting a method for transmitting and receiving a short coded message about 10 decimal digits in length over commercial telephone lines. The equipment should be simple and reliable, and capable of application to any commercial circuit without adjustment of the terminal equipment or the circuit itself. Transmission of digits should be at the highest speed consistent with reliable operation. In the following pages we consider the limitations imposed by the characteristics of telephone circuits and how they influence the choice of a signaling method. We then describe a reliable system which is compatible with these limita-

tions, and the results of tests of this system. Tests with some of the newer carrier systems are not included, since they were introduced after the completion of our work.

CHARACTERISTICS OF TELEPHONE CIRCUITS

Bell System circuits which are used for the transmission of speech comprise a variety of loaded, non-loaded and carrier circuits. The characteristics of these circuits differ widely, and while they do not impair speech transmission, they do affect the transmission of high-speed digital information. Circuits designed for special purposes, such as program transmission or telephoto, will not be considered.

Non-loaded voice-frequency circuits are ordinarily short, with attenuation increasing approximately as the square root of the frequency, and have no well defined upper cut-off frequency. As they present no serious obstacles to realizing our objectives, they will not be considered further.

Loaded voice-frequency circuits are used for both local and toll service, and have well defined upper cut-off frequencies. The attenuation characteristics of a number of typical circuits of this sort are shown in Fig. 1. In the case of local circuits, the upper cut-off frequency is set by the type of loading employed, and in toll circuits by the filters used in the four-wire terminating sets or two-wire repeaters. The terminating sets are designed to function as high-pass filters. Repeating coils also limit the low-frequency response. If we define the cut-off frequencies as the half-power points, or points at which the loss is 3 db, we see that a universal system is limited by these circuits to a band from about 300 to 2,000 cycles.

When we examine the attenuation characteristics of a number of typical carrier circuits shown in Fig. 2, we find the situation somewhat better at high frequencies and worse at low frequencies. The curves A-1 and A-2 are for the channel banks used in J, K, and L carrier systems. The approximate bandwidth between the limiting curves is from 400 to 2,600 cycles. The net band available for a universal system has now been reduced to 400 to 2,000 cycles. A further reduction to 400 to 1,700 cycles is necessary on two counts. The EB channel banks, or "emergency" circuits introduced during the war to provide two voice circuits over a single broadband carrier circuit have a band of from 400 to 1,700 cycles and an appreciable number of these circuits are still in use. Phase distortion, particularly at low frequencies, further limits the useful bandwidth even though EB circuits are excluded from consideration.

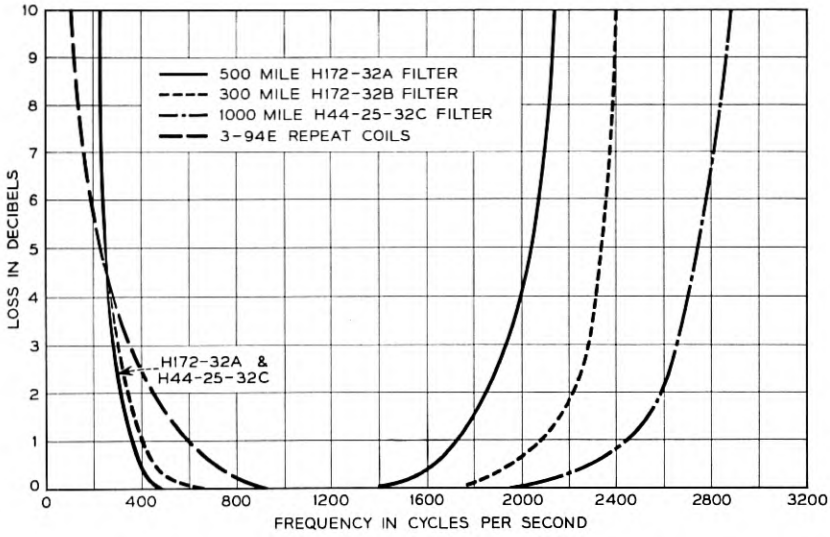


Fig. 1 — Transmission frequency characteristics of typical voice-frequency telephone circuits.

The delay distortion due to the non-linear phase characteristics of a number of typical limiting circuits is shown in Fig. 3. Since only the variation of delay with frequency is important, the curves have all been drawn to pass through the same point at a frequency of 1,000 cycles. The effect of this type of distortion depends upon the type of signal and the method of detection. The pulse distortion can be calculated,[1, 2] but it is difficult to relate pulse distortion to the performance of a particular system. We have experimentally determined that delay distortion about equal to 3 times the pulse length can be tolerated in a system of the type which we will describe below.* Inspection of Fig. 3 shows that for the delay distortion at the lower edge of the band to be equal to that at 1,700 cycles, we should fix the lower cut-off frequency at about 700 cycles. Application of the above rule of thumb indicates that pulses of about $3.9 \div 3 = 1.3$ milliseconds could be transmitted with tolerable delay distortion. This corresponds to a data rate of about 750 bits per second, which we have experimentally found to be optimistically high for reliable operation.

The majority of carrier circuits in the Bell System employ single-side-

* C. B. Feldman and A. C. Norwine have stated this result in an unpublished memorandum.

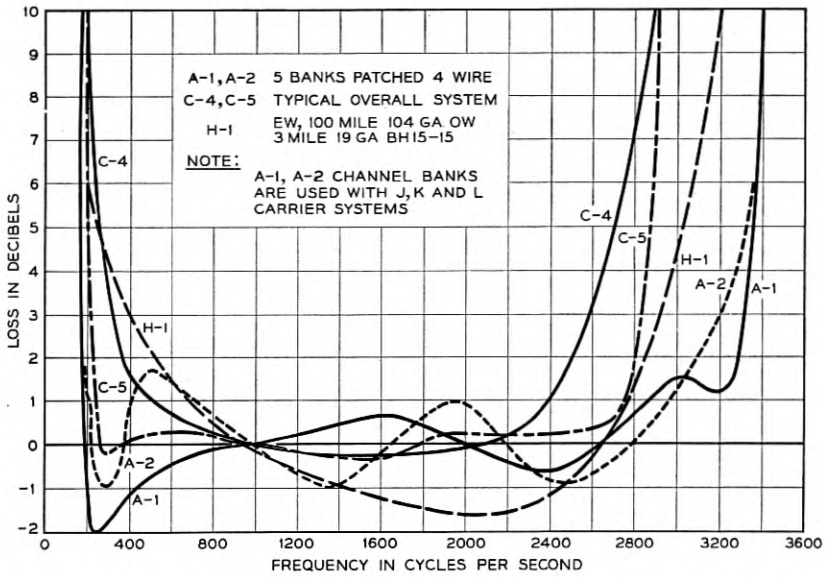


Fig. 2 — Transmission frequency characteristics of typical carrier telephone circuits.

band transmission with independent, non-synchronous carrier supplies for modulation and demodulation. The frequency difference is ordinarily only a few cycles per second, referred to the voice band, and has no noticeable effect on speech transmission. It is sufficiently great, however, to impose serious limitations on high-speed signaling. It can be shown that a signal $S(t)$ which is transmitted over such a circuit is received as

$$R(t) = S(t) \cos (\eta t + \epsilon) + Q(t) \sin (\eta t + \epsilon)$$

in which η and ϵ are respectively the radian frequency and phase difference between the two carriers. $Q(t)$ is the quadrature component which has the same amplitude spectrum as the original signal with the phases of all components shifted 90°. This shows that the structure of the received signal varies with time and this imposes another limitation on the type of signaling system which can be used. The envelope of the received signal is

$$\sqrt{S(t)^2 + Q(t)^2}$$

which is not a replica of the original signal, but is invariant with time.

Other limitations are imposed by noise, crosstalk, and echoes. The

CURVE	1000 ∿ DELAY	DESCRIPTION
1	58.64	3000 MILES 5 LINK K2 SYSTEM CHANNEL 12 PLUS 500 MILES OF 19-H-44 PLUS 6 PAIRS OF 4W TERMINAL SETS
2	53.30	1000 MILES OF 19-H-44 PLUS ONE PAIR OF 4W TERMINAL SETS
3	12.26	2000 MILES 3 LINK J2 SYSTEM CHANNEL 1 W-E JSA PLUS 3 PAIRS OF 4W TERMINAL SETS
4	22.70	200 MILES OF 19-H-174 PLUS ONE PAIR OF 4W TERMINAL SETS

TWO-WIRE VF PATCHING IS ASSUMED

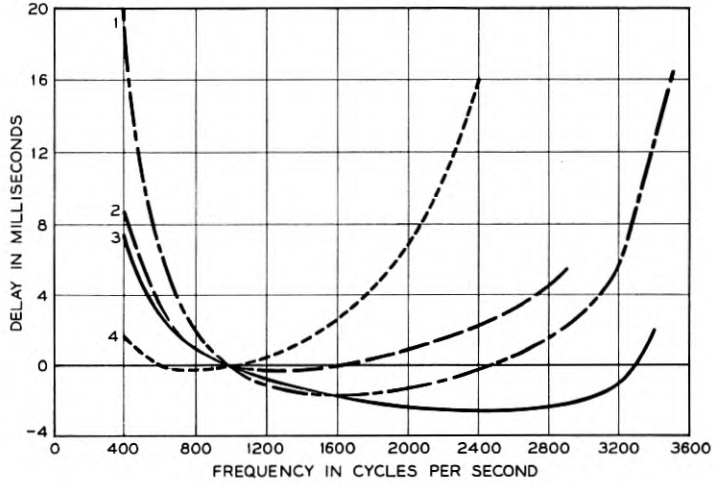


Fig. 3 — Delay distortion of typical maximum length telephone circuits.

amounts of the latter two are ordinarily small and have substantially no effect on the type of signaling system which will be described. They may be a factor in other, more vulnerable systems. We have found no adequate theoretical way of dealing with the effects of noise in a quantitative manner. In the case of thermal noise, an approximate treatment is possible, but departures from the idealized mathematical model render this treatment only approximate. The subject is best studied experimentally and will be discussed under "Results."

Another possible limitation is imposed by the use of compandors on some circuits. These devices compress the amplitude range of the signal at or near the transmitting end of the circuit and expand it, or restore it to normal, at the receiving end of the circuit. Multi-level signals

might be expected to be vulnerable to this kind of treatment. The effects of compandors on relatively noise-free circuits have been studied experimentally and will be discussed under "Results."

CHOICE OF A SYSTEM

We have seen that the pass-bands of telephone systems restrict us to the use of signals whose frequency spectra lie between about 400 and 1,700 cycles, and that delay distortion considerations require that the lower limit be raised to about 700 cycles. This immediately rules out of consideration all methods of signaling which require transmission of a dc component. The obvious choice is to use a method which involves one or more modulated carrier frequencies. This is also attractive because envelope detection can be readily employed to avoid the difficulties encountered with non-synchronous carriers. Other methods which do not depend upon the precise structure of the signal within the pulse interval are possible, and have been considered both theoretically and experimentally. Such systems are generally more complex and no more reliable in the presence of noise than simple envelope detection of a modulated carrier.

The multifrequency signaling system[3] now in use for inter-toll signaling employs six carrier frequencies, in a two-out-of-five self-checking code. (Five of the frequencies are used to represent the ten decimal digits, the sixth frequency is required for supplementary signals.) For a given available frequency band, the rate at which information can be transmitted is in principle independent of the number of carrier frequencies employed. Actually there will be some reduction in speed with a multifrequency system due to wasted frequency space between bands. The amount of equipment required at the transmitting and receiving terminals increases with the number of carrier frequencies employed.

The effects of delay distortion are minimized in a multichannel system, such as voice-frequency telegraph, in which the available frequency band is divided into a number of bands each having substantially less distortion than the entire band. The average total delay in each of these narrow bands will differ from one band to another, and various parts of the message will be received at times which will differ with each type of circuit employed. In some types of systems, this may be of no importance, but generally speaking, it introduces additional complications in the equipment.

When signal elements consist of more than a single carrier frequency present at one time, care must be taken to avoid transient interference

between channels. This can be adequately controlled by the use of transmitting filters, and timing in the receiver to inspect the signal after the transient has decayed to a negligible amplitude.

A system of this sort is vulnerable to "twist," or differences in attenuation of the different frequencies. Some sort of automatic volume control is required in the signal receiver, and some types accentuate the difference in amplitude of two or more frequencies applied to their inputs.

We have experimentally studied some multifrequency signaling systems and have found that they are somewhat more vulnerable to noise than a single frequency system. Multichannel systems are less vulnerable to some types of noise, but both multifrequency and multichannel systems are more complex equipment-wise than single frequency systems. We have been led to the conclusion that a single-frequency system employing envelope detection is to be preferred for high-speed systems.

We must now consider the relative merits of single sideband versus double sideband transmission. At first glance, we might consider that a single sideband system could transmit at twice the rate possible with a double sideband system. Since complete separation of sidebands is not possible when the base-band signal has a dc component it is necessary to transmit a vestigial sideband which occupies an appreciable portion of the available transmission band. Because of the effects of non-synchronous carrier systems, it is necessary that the transmitted signal be accompanied by some non-modulated carrier which is present at all times during signaling and which carries no information. Since the maximum power which can be applied to telephone circuits is limited, this results in forcing a reduction in the power of the information bearing signals, with a corresponding increase in vulnerability to noise. A vestigial sideband system requires a carefully designed filter with linear phase through the cut-off frequency as it is sensitive to delay distortion in the neighborhood of the carrier frequency.

A double sideband system requires no carrier during absence of signal, and because of the redundancy of the two sidebands, is less sensitive to amplitude and phase distortion. Considering the important case of negligible amplitude distortion, it has been shown[4] that the effective phase shift of the recovered signal is the average of the phase shift in each sideband or

$$\psi = \frac{(\varphi_+ - \varphi) + (\varphi - \varphi_-)}{2}$$

in which φ is the phase shift at the carrier frequency and φ_+ and φ_- the phase shifts in the upper and lower sidebands.

Although we have not attempted a quantitative evaluation of the relative merits of the two methods of transmission and have not made a direct experimental comparison between the two, the above considerations have led us to conclude that the double sideband system is best adapted to meet our objectives. We will restrict further discussion to such systems.

It is generally accepted that the maximum reliability in the presence of noise is obtained when a code is employed which consists of signal elements which are either present or absent. Higher signaling speeds can be obtained at the expense of noise impairment if multi-amplitude codes are employed. Since we are concerned with maximum reliability we consider only on-off types of signals.

When the complete message is short, the advantages of one code over another are not great, and the particular choice of code is usually fixed by the general objectives in view, or by equipment considerations. When information is received serially, it can be transmitted substantially as received, whereas if the message is coded as a unit, a delay equal to the length of the message is introduced in the coding process and again in the decoding process. We have elected to encode our message by decimal digits in two-out-of-five form to take advantage of the self-checking feature of this type of code.

It seems advisable to review some of the relations which must exist between carrier and sidebands if distortion is to be avoided. The carrier frequency must be higher than the highest frequency component of the base-band signal, and components of the base-band signal must not intrude upon the lower sideband. The base-band signal can be eliminated by balance in the modulator or by filtering if it is first modulated with a high-frequency carrier and then located at the desired point in the transmission band by a second process of modulation. If the carrier frequency is higher than twice the highest base-band frequency, there is no overlap and the base-band signal can be eliminated by a high pass filter. It is fortunate that the location and width of the available transmission band permit us to use the last method.

The maximum theoretical speed is obtained by postulating a signal which has a standard value at a sample point, say at the center of the signal interval and which is zero at the sample point of all other signal intervals. Since the locations of the zeroes are dependent on delay distortion, and since this differs from circuit to circuit, we cannot depend on this property to recognize the presence of a signal and a reduction in repetition rate is required. Even if the maximum speed could be

obtained, the presence of noise would require that a lower rate be employed to insure reliable operation.

Ample operating margins are obtained with a pulse repetition rate equal to half the maximum theoretical rate. Higher rates can be used with increased vulnerability to noise. We believe that performance will be intolerably degraded if the repetition rate exceeds three-fourths of the maximum theoretical rate.

To minimize the effects of delay distortion, the frequency spectrum of the signal should be confined to the band in which the distortion is low. To minimize oscillations at the edges of a pulse, the cut-off should be gradual. These requirements are met by a Gaussian base-band pulse, since the frequency spectrum of such a pulse is also Gaussian. A single section low-pass filter with a nominal cut-off frequency equal to the reciprocal of twice the width of a rectangular pulse has a response to such a pulse which is a fair approximation to a Gaussian curve as shown in Fig. 4. A simple filter of this sort has been found satisfactory for this purpose. Somewhat better results have been obtained with empirically optimized pulse shapes generated by a capacitive commutator which is discussed below.

SYSTEM DESCRIPTION

A system which operates in accordance with the principles outlined above can be mechanized in many ways. One version, which was used in our experimental work, is shown in Fig. 5. The transmitter and receiver are connected to the ends of the chosen transmission facility, and the message is transmitted repetitively on a start-stop basis until a two-out-of-five check indicates that a plausible message has been received. The receiver then returns a stop signal to the transmitter and both are disconnected from the facility.

The message consists of a start and synchronize signal followed by an eight digit number, each digit of which is coded in two-out-of-five form. The number to be transmitted consists of 40 bits and is stored in a register in parallel form. It is removed from the register in serial form by a scanner and the pulses are shaped by a low-pass filter and applied to a modulator to produce double sideband carrier pulses for transmission over the line. The scanner shown in Fig. 5 consists of germanium diode gates controlled by vacuum tube ring counters.

As an alternative, a capacitive commutator was used to perform the

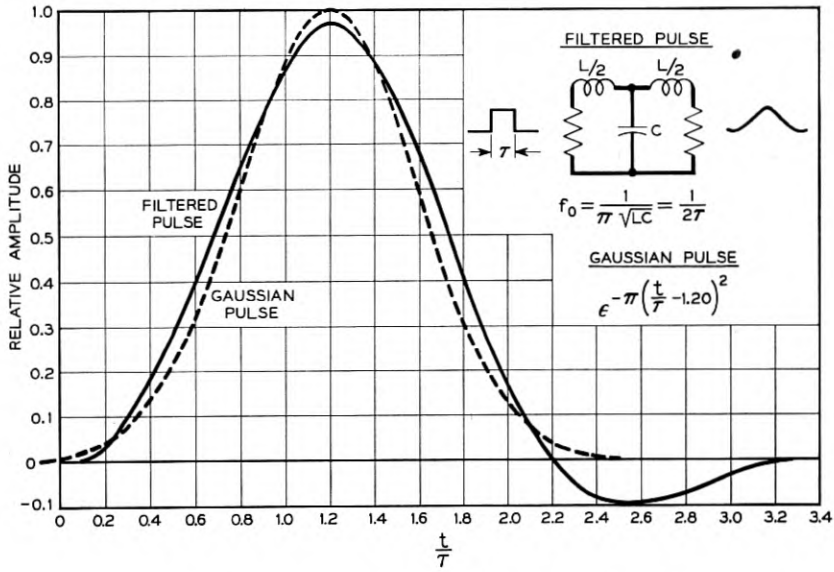


Fig. 4 — Filtered rectangular pulse and Gaussian pulse.

functions of the scanner, low-pass filter, and modulator. This device consists of two glass discs about three inches in diameter, one stationary and the other rotated just out of contact. Carrier frequency is applied to conducting segments, painted on the stationary disc, in accordance with the coded signal. The signals appear in sequence on one conducting segment on the rotating plate, which is coupled to the stationary plate with capacitive slip rings. The shape of the resulting pulses is determined by the size and shape of the scanning segment.

The received signals enter an automatic volume control circuit which amplifies and delivers signals to an envelope detector at substantially constant amplitude, over a 30-db range of inputs. The detected signal is then filtered and sliced. The start circuit recognizes the start signal and starts a start-stop oscillator upon the arrival of the synchronize pulse. The digit pulses are directed to the register by means which involve coincidence between sliced signal pulses, bit sampling pulses, and digit gating pulses.

We have employed vacuum tubes and germanium diodes as the basic circuit elements. There are many well known methods of implementing the various functions with these or other components, and consequently, we will omit detailed circuit considerations.

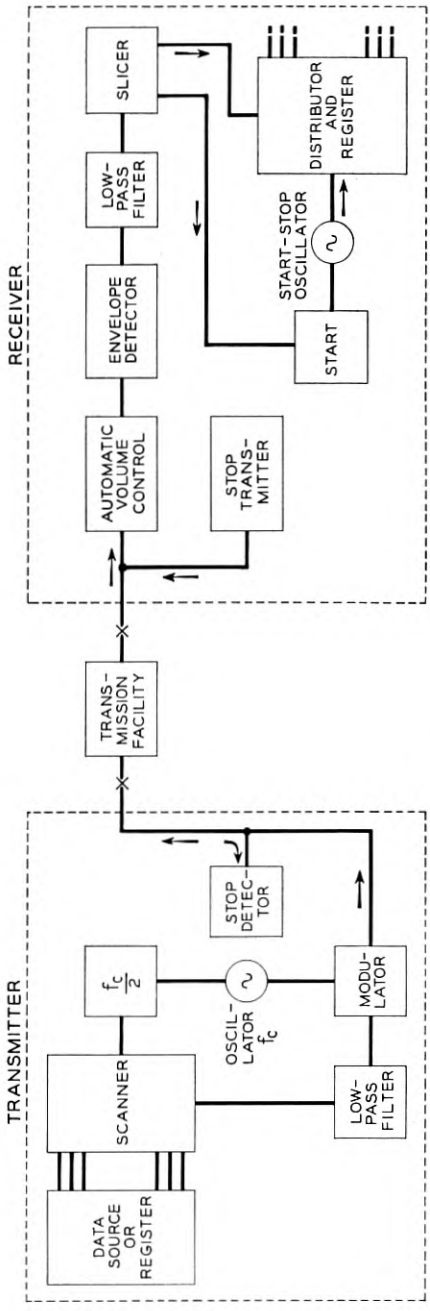


Fig. 5 — Block diagram of signaling system.

TESTS

The system has been tested in the laboratory and on commercial toll circuits in the Long Lines network. The test set-up is shown in Fig. 6. A number selected for transmission was manually set in a comparator and compared with the number actually received. If the two numbers agreed, an "OK" call was counted. If the two-out-of-five check indicated an implausible received number, a re-cycle was recorded and the transmission was repeated until a check was finally obtained. If a plausible but incorrect number was received, an error was recorded. A call was considered mutilated if either a re-cycle or an error was recorded. A number was transmitted at least 100 times for each test condition, and many more observations were made when the mutilation rate was low.

Laboratory tests were made with facilities which included artificial non-loaded lines, carrier terminals, and compandors. One of the carrier terminals was modified so that the frequency difference between carriers at the modulator and demodulator could be varied continuously from 0 to 20 cycles per second. Various band-pass filters were inserted in the transmission facility. Tests were also made over representative commercial telephone circuits and over a number of such circuits in cascade.

The effects of noise were studied by introducing different types of interference through a hybrid circuit. The majority of these tests were made with thermal noise but other tests were made with pulsed thermal noise, contact noise and single frequency tones.

We shall not attempt to characterize all the different types of interference which we employed. The properties of thermal noise are well known and need no description. Since impulse noise has not been studied

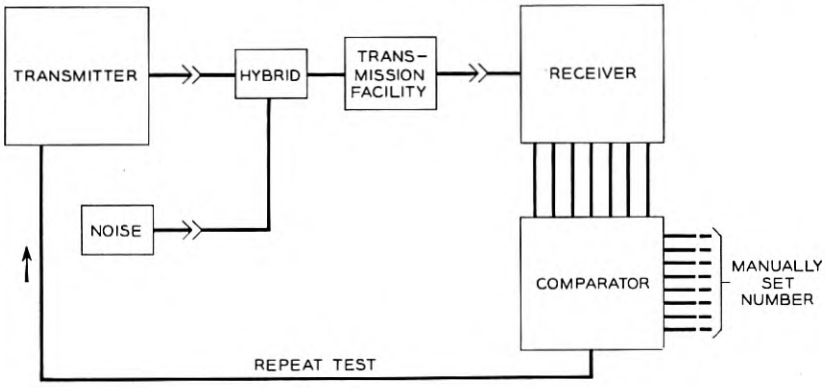


Fig. 6 — Block diagram of test set-up.

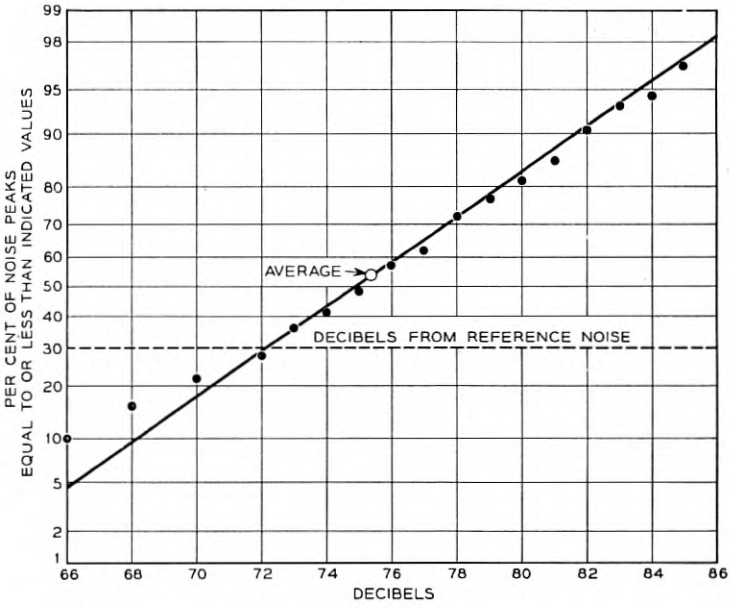


Fig. 7 — Distribution of recorded contact noise peaks. Measured with a 2B noise meter.

extensively, we will discuss the properties of a single sample of recorded contact noise which was employed in our test.

The noise peak voltages were measured with a recording peak voltmeter which could resolve peaks with durations of the order of a millisecond or less. Measurements were also made with a 2B noise meter which is a standard Bell System instrument.[5] Because of the relatively long time constant of this meter, a short peak is measured with an apparent amplitude which is much less than the true amplitude, and this difference will depend upon the duration of the peak. Comparison of the peak voltage measurements with those made with the 2B noise meter showed a reasonable correlation between the two distributions, with a correlation coefficient of 0.78. A short noise peak equal in amplitude to the amplitude of a 0 dbm sine wave is measured by the 2B set as approximately 66 db above reference noise, or on the average the 2B set will measure short noise peaks as if they were about 24 db less than their true amplitude.

RESULTS

Preliminary tests were conducted over representative toll circuits to determine the optimum location of the carrier frequency. This depends

upon the type of circuit, but the best compromise appeared to be in the neighborhood of 1,200 cycles. With this choice optimum results were obtained with sidebands symmetrically located about the carrier frequency and occupying the band from 700 to 1,700 cycles. Both 1,000- and 1,200-cycle carrier frequencies were employed in many of our tests.

Laboratory tests using representative lengths of artificial non-loaded cable showed adequate margins. The maximum "twist" in the signaling band was 10 db. No attempt was made to determine the maximum twist that could be tolerated.

From one to four "K" carrier links caused no observable degradation in signaling performance. Frequency differences between the carriers of up to 20 cycles, and continuous variations, did not affect performance at signaling speeds of 100 to 150 digits per second, or 500 to 600 bits per second.

From one to four compandors were inserted in a relatively noise free transmission path with no effect on reliable transmission of pulses. We believe that two properties of the messages contribute to this result. The start pulse, which is long compared to a single bit, prepares the compandor for the signal bits, which follow it. The compandor is then held at a more or less fixed setting, since the use of a two-out-of-five code insures that there will be no long interval between pulses.

Recordings of babble or crosstalk between channels of EB carrier circuits were used to show that cases of this type of interference, which were bad by telephone standards, had no effect on reliable signaling.

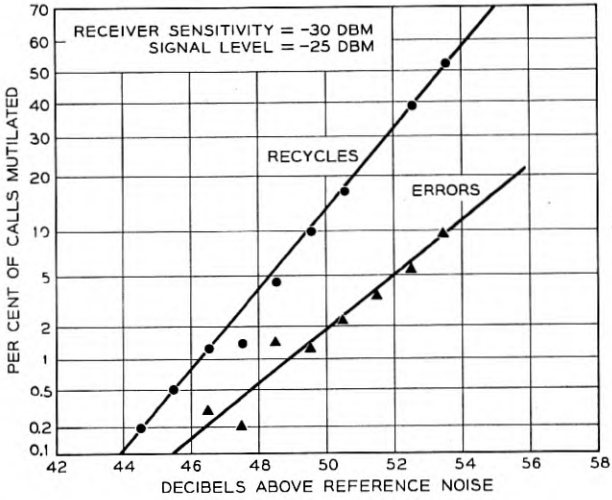


Fig. 8 — Per cent of mutilated calls versus thermal noise power.

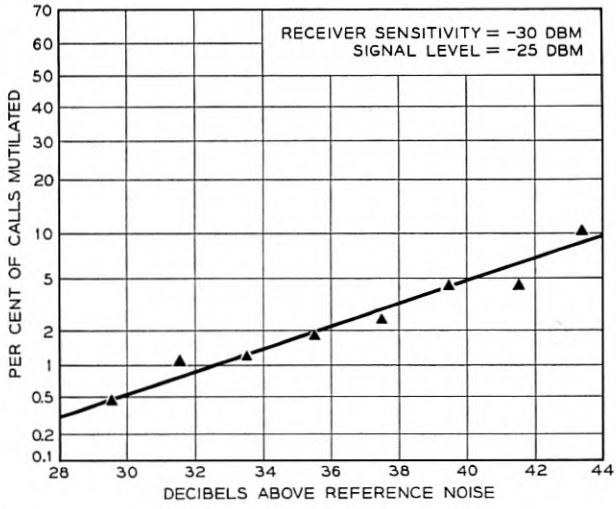


Fig. 9 — Per cent of mutilated calls versus recorded contact noise power.

The effect of thermal noise on the reliability of signaling is shown in Fig. 8. In these tests the peak of the received signal was equal to the peak of a −25 dbm sine wave. The receiver was set to just operate on a 5 db weaker signal, or −30 dbm. A 500 to 1,500 band-pass filter was ahead of the receiver and the carrier frequency was 1,000 cycles. An eight digit number was transmitted but mutilations were recorded for only the first three digits. It is interesting to note that the self-checking code provides a substantial but not perfect check on mutilated signals. Mutilation of two or more of the bits representing a decimal digit causes the failures. The effect of noise at a low level is ordinarily to add a false signal element, or to nullify a true one. Either of these will be detected by the self-checking feature. At higher noise levels there is an increasing tendency to have both effects occur within a single digit. Since the mutilated digits contain two signal elements, one of which is incorrect, the self-checking circuits will fail to detect the error. Since Bell System Standards require that message circuit noise shall be less than 26 db from reference noise at the zero level point, it is clear that the system has ample margin against thermal noise.

Similar data for impulse noise having the peak distribution shown on Fig. 7 are shown in Fig. 9, with errors and re-cycles combined into a single curve. The noise values shown are taken as the value at the 30 per cent point of the distribution curve as indicated above. The margin against noise mutilation for this type of noise is much less than for

thermal noise, but reasonably satisfactory performance for telephone signaling is to be expected when circuit noise meets Bell System Standards.

In the case of thermal noise, the maximum peak voltage that we have observed is about five times the r.m.s. voltage, or a difference of 14 db. R.m.s. noise 14 db below the just operate point of the receiver, $-30 - 14 = -44$ dbm should rarely cause a mutilation. Since a sine wave at -90 dbm will indicate reference noise on a 2B noise meter, the tolerable amount of noise is approximately $90 - 44 = 46$ db above reference noise. This figure is consistent with the data of Fig. 8.

Peak voltage measurements of the recorded contact noise show that the actual peaks are about 24 db greater than those measured by the 2B set. We would expect satisfactory operation with impulsive noise about 24 db lower than with thermal noise, or $46 - 24 = 22$ db above reference noise, which is in approximate agreement with Fig. 9.

The system was tested on nine representative toll circuits and four circuits composed of several voice circuits in tandem to produce conditions of far greater severity than would ever be encountered on com-

TABLE I — SUMMARY OF TOLL LINE TESTS

Test	Circuit Loop	Miles Approx.	Facilities	Per cent Re-cylces			
				100 D.P.S.		150 D.P.S.	
				Cap Com	Elec	Cap Com	Elec
1	NY-St. Louis	1,880	"EB" A on "K" Carrier	0	0	0	0
2	NY-St. Louis	1,880	"EB" B on "K" Carrier	0	0	0	6.2
3	1 + 2	3,760	1 + 2	0	0	0	7.6
4	NY-Knoxville	890	"K" + "C" Carrier +4W 19-H-44	0.2	0	0	11.7
5	NY-Jacksonville	2,010	"K" + "L" Carrier	0	0	0	8.2
6	NY-San Francisco	5,950	4 "K" Carrier 2 "K + L" + "L" Carrier	0	0	0	0
7	NY-San Francisco	5,950	"EB" on "K" Carrier "A" and "B" Banks	0	0	0	0.2
8	6 + 7	11,900	6 + 7	0	67	4.2	51
9	NY-Harrisburg	355	4W 19-H-25	0	0	0	.7
10	8 + 9	12,255	8 + 9	0	47	0.2	N.O.
11	6 + 9	6,305	6 + 9	0	0	0	26*
12	NY-Albany	310	2W 16-H-44	0	0	0	0
13	NY-New Haven	156	2W 19-H-174	0	0.3	0	0

N.O.—This condition not operable.
* One error recorded on this condition.

TABLE II — SUMMARY OF OPERATING RANGES IN TOLL LINE TESTS

Test	Range, db in Attenuator			
	100 D.P.S.		150 D.P.S.	
	Cap Com	Elec	Cap Com	Elec
1	0–29	0–31.5	0–26.5	1–24
2	1–32	0–33	4–31	26–32
3	0–27	0–28	4–26	26–27
4	0–31	0–32	0–31	22–30
5	0–33	0–33	0–32	6–31
6	0–33	3–32	5–32	31–33
7	2–35	3–36	4–33	17–34
8	4–35	28	15–34	35
9	0–25	0–27	0–24	0–23
10	0–30	0–31	5–29	*
11	0–26	0–27	1–21	25–26
12	0–25	0–26	0–23	0–24
13	0–14	0–16	0–13	0–15

* In this case circuit was inoperable. In some cases above where one limit is zero, it is probably that the range was considerably greater than indicated above since the normal operating range at 100 digits per second is about 32 db.

mercial circuits. "N" and "O" carrier were not in use at the time of the tests.

Tests were made at two speeds, 100 decimal digits per second and 150 decimal digits per second with an electronic transmitter and with a capacitive scanner. A 500-cycle low-pass filter was used with the electronic transmitter at both signaling speeds. No filter was required with the capacitive commutator. An attenuator preceding the receiver was varied to determine the range of attenuation over which the system would operate satisfactorily. The results are shown in Tables I and II. These tests indicate that a signaling speed of 750 bits per second or higher might be realized. We consider this to be an optimistic conclusion because of the favorable noise conditions during the tests. We believe that signals can be transmitted reliably at about 650 bits per second.

CONCLUSIONS

One of the fundamental limitations on pulse signaling is the phase distortion inherent on all commercial telephone circuits. Without phase equalization and with signals transmitted in a single band, the practical location of this band is from about 700 to 1,700 cycles. Methods of signaling which depend upon recognition of the precise structure of the signal are not practical because of the variable phase shift introduced by non-synchronous carriers of most carrier systems. The envelope of a

signal modulated carrier is not dependent upon this type of phase shift. A signaling system which transmits pulses of a 1,200-cycle carrier frequency and employs envelope detection has been found to be reliable at signaling speeds of about 650 bits per second. This system operates with adequate margin over substantially all commercial telephone circuits that were in use at the time the tests were made.

References

1. H. A. Wheeler, The Interpretation of Amplitude and Phase Distortion in Terms of Paired Echoes, Proc. I.R.E., June, 1939.
2. E. D. Sunde, Theoretical Fundamentals of Pulse Transmission, B.S.T.J., May and July, 1954.
3. C. A. Dahlbom, A. W. Horton, Jr. and D. L. Moody, Applications of Multifrequency Pulsing in Switching, Trans. A.I.E.E., **68**, pp. 392–396, 1949.
4. R. V. L. Hartley, Relations of Carrier and Side-Bands in Radio Transmission, B.S.T.J., April, 1923.
. P. Mertz, Transmission Line Characteristics and Effects on Pulse Transmission, Symposium on Information Networks, Microwave Research Institute, New York, April 13, 1954.
J. V. Harrington, P. Rosen and D. A. Spaeth, Some Results on the Transmission of Pulses over Telephone Lines, Symposium on Information Networks, Microwave Research Institute, New York, April 13, 1954.

Experimental Extrusion of Aluminum Cable Sheath at Bell Telephone Laboratories

By G. M. BOUTON, J. H. HEISS, and G. S. PHIPPS

New techniques for extruding aluminum directly over paper insulated cable core at low temperature and pressure are described. Pistons operate from opposite ends of a cylinder to force the aluminum through a circular orifice between a die and core-tube located on an axis perpendicular to that of the pistons and midway between them. The inter-relation of extrusion temperature, pressure and rate, as well as the quality of the sheath produced, are found to be dependent on the die and core-tube design. Special lubrication and press charging techniques are discussed.

INTRODUCTION

In recent years considerable attention has been focussed on the possibility of using aluminum instead of lead as cable sheath. Over fifteen years ago German publications[1, 2] described activities abroad. It was emphasized at that time that for a given temperature, aluminum would require much higher extrusion pressures than lead. Also, the pressure required for extrusion of commercially pure aluminum (99 per cent) was reported to be considerably higher than that for super purity aluminum (99.99 per cent) under a given set of conditions. The possibility of obtaining lower pressures by raising the temperature is limited by the danger of scorching the paper or other organic insulation of the core. It is evident, therefore, that improvements in extrusion techniques are required. Progress in this respect has been reported recently.[3, 4, 5, 6]

Lower cost is an important part of the incentive for using aluminum instead of other types of sheath. In the telephone industry lead has already been supplanted to a sizable extent by composite sheaths, such as alpeth or stalpeth. The former consists of a thin corrugated aluminum sheath covered with extruded polyethylene-carbon black compound. The overlap seam in the aluminum is sealed with an organic adhesive. Stalpeth has a thin corrugated aluminum inner sheath without overlap or seal. Outside of this there is a sheath of corrugated tinned, steel sheet

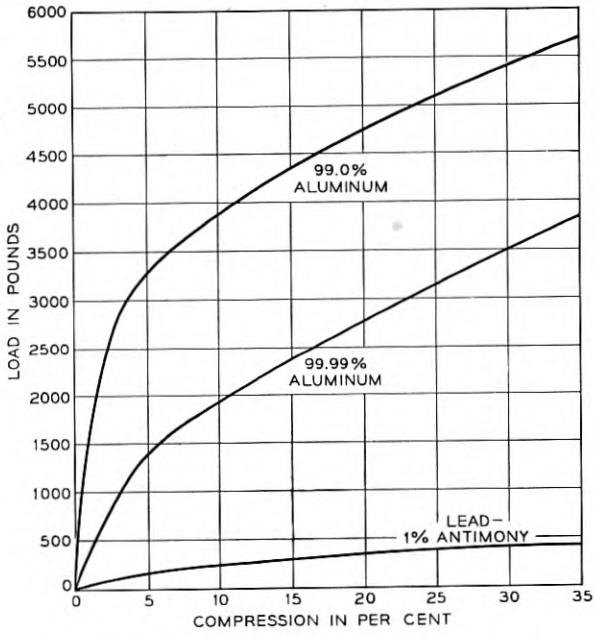


Fig. 1 — Compression tests at 265°C.

with a soldered seam. The whole is covered with extruded polyethylene compound. An all aluminum sheath would have a cost advantage over all types of sheath discussed. In addition, the aluminum sheath would be lighter and have greater tensile and creep strength.

Fatigue and corrosion must be explored further. The fatigue limit of aluminum at 10^8 cycles is considerably above that of lead on a stress basis. However, in aerial service, the strains might be similar in lead and aluminum sheaths since their temperature coefficients of expansion are similar and the copper core exerts a controlling effect on both types of sheath. Because of the higher modulus of elasticity of aluminum, a given strain results in a much higher stress than in lead. It is known that temperature fluctuations cause strains in aerial cable sheath considerably above the fatigue endurance limit. Until strain data are collected on sheath in service environments and fatigue tests are conducted at such strains, the field life expectancy of aluminum sheathed cables will be subject to question.

EXPLORATORY STUDIES

Plasticity studies were made at an early stage of the aluminum sheath

investigation. The method used had proved quite successful in estimating extrusion pressure characteristics of lead alloys.[7] It consisted in compressing cylinders 0.92″ in diameter and 1.5″ long at a fixed temperature and a constant rate of 0.1″ per minute between parallel plates and plotting the load corresponding to the deformation. Fig. 1 shows some typical results obtained in this manner. By comparing the loads to produce a given deformation above the yield point, the relative extrudability may be estimated. The data indicate that about ten times as much pressure would be required to extrude 99.99 per cent aluminum compared to lead-1 per cent antimony at 265°C. The estimated pressure would be even greater for less pure aluminum. Since some 50,000 psi are required to extrude this widely used lead cable sheathing alloy through conventional extrusion passages, it is evident that the pressures required for aluminum extrusion would be far beyond the strength of available structural materials.

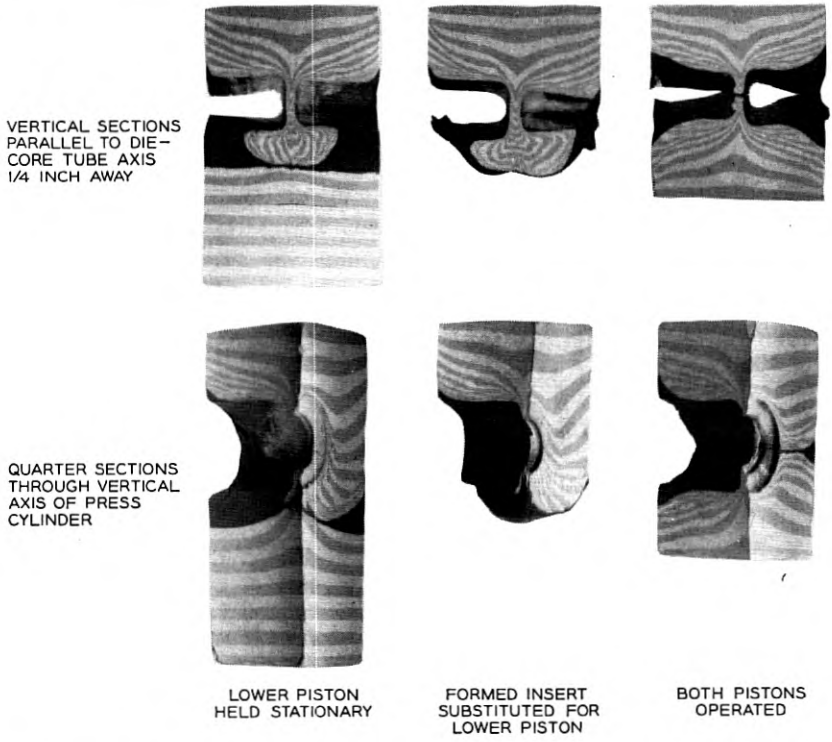


Fig. 2 — Flow patterns seen in sections cut through laminated modeling clay cylinder residues after extrusion.

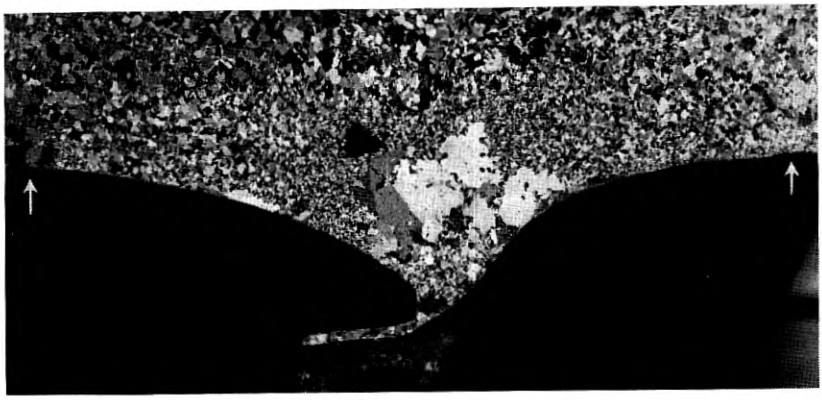


Fig. 3 — Section through a cylinder residue removed from the extrusion press. Arrows indicate the residual black layer of lead-0.1 per cent antimony from the original charge. Several subsequent charges of pure lead failed to remove the antimonial layer from the surface of the die and core-tube even close to the point of sheath formation near the middle of the picture.

With these considerations in mind, new designs of extrusion equipment were studied. An early experimental press was constructed with two axially opposed pistons and the die and core-tube located midway between them at 90° to the cylinder bore. Valuable information on flow patterns was obtained using colored modeling clay laminates. Fig. 2 shows typical results of these experiments. Early studies were made by fixing one piston and examining the flow of the laminates through an unlaminated different colored mass in the die and core tube region. This gave some clue as to what an ideal cavity would be for a one-piston extrusion where the clay was free to choose its own flow channels. Considerable volumes of the dark undisplaced clay can be seen in both sections in the first column of Fig. 2. By cutting away the stationary clay, the preferred shape of the shear surfaces in the central region could be obtained. A plaster mold was made to this shape and used in place of the bottom piston. Additional extrusion of laminates established that there was always a stationary layer of modeling clay next to the die-block surfaces as shown in the second column of Fig. 2. Flow took place by shear within the clay rather than by flow along the surfaces that were supposed to guide it. The third column shows flow when both pistons were operated. Here again much of the dark colored clay remained on the surface of the tools after several extrusions.

This stationary layer was further confirmed by experiments with lead. Incorporation of about 0.1 per cent antimony does not seriously alter

the extrusion characteristics of pure lead. When this alloy is etched, it has a black matte appearance readily distinguished from that of the surrounding lead. Fig. 3 shows the results of extrusions using both pistons where first lead-0.1 per cent antimony was charged followed by several charges of pure lead. The etched residue shows the black adherent layer adjacent to the cavities left by the die and core-tube.

On the basis of these studies, attempts to develop a preferred die-block contour were abandoned. Instead, the chamber in the region of the extrusion orifice was left as open as possible. This permits the flowing material to choose its preferred path, which, indeed, may change slightly for different temperatures, extrusion rates and materials.

EXPERIMENTAL EXTRUSION PRESS DESIGN

As a result of this work a new experimental press was designed to permit additional study of basic principles of the plastic flow of metals

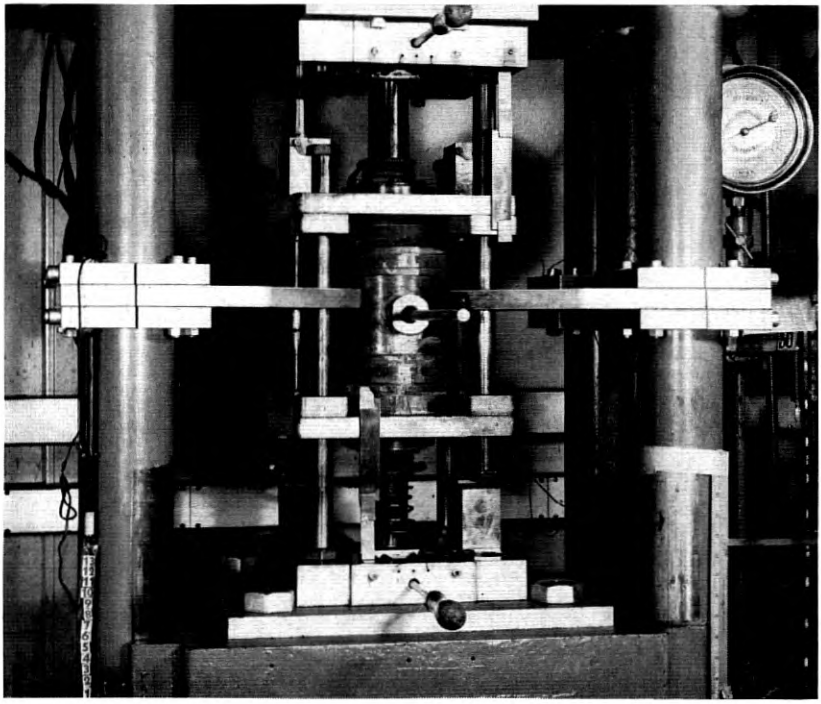


Fig. 4 — Front view of experimental extrusion equipment installed in press.

through chambers and extrusion dies. The pressure source is a single hydraulic cylinder press of 180-ton capacity. The extrusion equipment has two pistons operating from opposite ends of the extrusion cylinder. The ratio of bore to stroke was made as large as possible because considerable pressure is lost through friction and shear of aluminum at cylinder walls. The use of two pistons operating from opposite ends of the cylinder insured a reasonable amount of material extruded per stroke. Two cylindrical aluminum billets measuring 2.2 inches in diameter and approximately 3.5 inches in length produce about 50 feet of cable sheath, one-half inch in diameter with an 0.028-inch wall. Fig. 4 shows the over-all picture of the extrusion equipment installed in the hydraulic press.

A schematic side view of the press showing the relative positions of the members is shown in Fig. 5. The cylinder block is constructed of three concentric shell tubings heat-shrunk together to provide a fairly high residual compressive stress on the inner wall of the cylinder. They

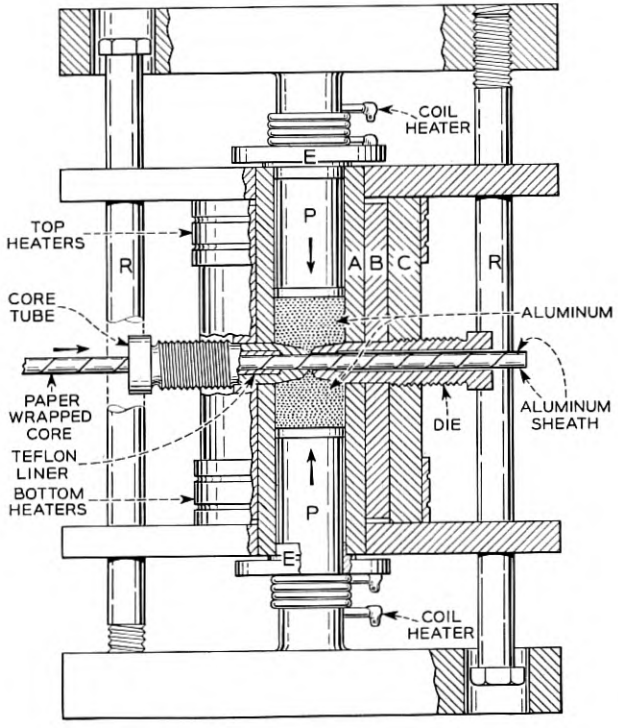


Fig. 5 — Diagram of experimental press assembly.

are shown schematically by the cross-hatched areas A, B, and C. The block is held in place by end plates having porous bronze bearings which slide on guide rods, R, and serve to keep the freely floating press in fairly exact alignment. When the press is opened, the pistons, P, slide back in guides, not shown, to permit introduction of fresh aluminum cylindrical billets into the extrusion cylinder. Sliding bronze bushings, E, attached to the pistons fit into the top and bottom of the cylinder and aid in maintaining alignment of the pistons in the cylinder during the application of extrusion pressure. The die and core-tube are shown in position at the middle of the cylinder. Their distance from each other determines the wall thickness of the sheath. Also pictured schematically is the formation of the cable sheath near the axis of the cylinder. The paper wrapped core is introduced from the left hand side and is carried through the press to the right by the extruding sheath. The core tube has been recessed over a large portion of the guide hole in order to accommodate a teflon liner. This plastic liner acts as a heat insulator which protects the paper wrap from being seriously scorched, especially during a recharge period. Heat is fed to the ends of the cylinder by means of the two sets of band heaters which are regulated by a recording controller. The pistons are heated by sliding coil heaters also shown in position. These extra heaters serve to balance the heat lost by conduction through the pistons and thus tend to provide fairly good temperature distribution along the cylinder wall.

The press may be operated either at constant applied pressure or constant extrusion rate. To operate at constant pressure the valves are manually controlled. A sliding brush attached to the platen of the press makes contact with a slide wire to indicate each ½-inch platen travel by sounding a buzzer and marking an indication on the recorder chart. Each time the buzzer sounds, a mark is placed on the sheath. This designates specimen lengths which are numbered 0 to 13, inclusive, over one full press stroke. Thus, matching the appropriate time from the recorder chart, the rate of extrusion may be obtained for any portion of the piston stroke.

The other method of operation, at constant extrusion rate, consists in regulating the rate of travel of the platen of the press. This is accomplished manually by making an indicator attached to the platen coincide with markings on a vertical tape which can be set to travel at a fixed speed. A recording pressure gauge coupled to the moving platen gives the pressure continuously during the piston stroke.

The location of the die and core-tube offers many advantages. The opposing operation of the pistons minimizes the bending moment on

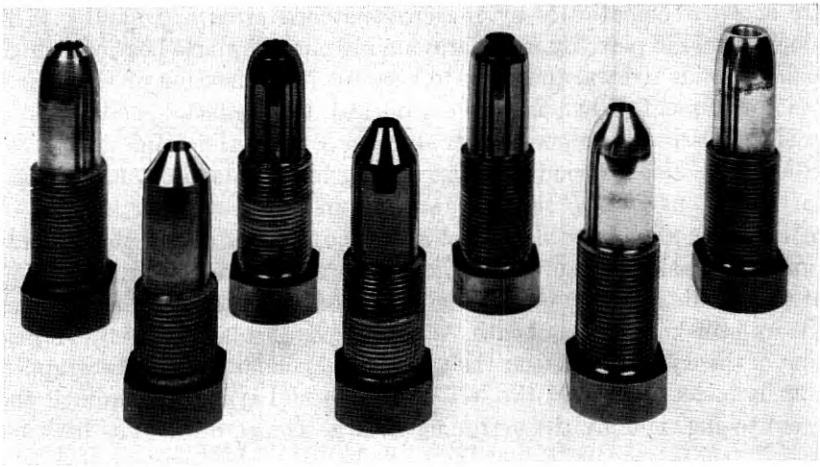


Fig. 6 — Some experimental dies and core-tubes. The three middle tools have baked on oil finish.

the tools and also greatly reduces the shear surfaces within the aluminum as it approaches the extrusion orifice. This construction results in far less dissipation of pressure than the conventional single piston lead presses in which some of the material from the top must flow completely around the core-tube, join and then flow through a highly restricted path to the point of sheath formation. Locating the tips of die and core-tube at the axis of the extrusion cylinder further reduces the shear surfaces over which the aluminum must move. Another advantage of this axial position is that the core-tube has greater rigidity because of its shorter cantilever length. There is a slight disadvantage in the placement of the die tip in the center of the cylinder because of greater bending moment, but this is greatly outweighed by having both the die and core-tube move in unison under any fortuitous pressure unbalance between the pistons thus helping to maintain sheath concentricity.

EFFECT OF DIE SHAPES

Early in the program of extrusion studies it became evident that the shape of the tip of both the die and core-tube would have an important bearing on the quality of the sheath and the extrusion rate. Fig. 6 shows a few of the dies and core-tubes used. Fig. 7 shows details of the shapes of the tips of some of the dies that were studied while Fig. 8 shows shapes of core-tubes used in various combinations with these dies. Each letter

shows a separate step in the modification of these tools. Early indications were that a sharp edge on the die (as in 8A, Fig. 7) resulted in high extrusion speeds, but in most circumstances the sheath lacked both roundness and smoothness, for example, the left-hand tube in Fig. 9. To correct for this condition the first studies involved the introduction of various radii at the extrusion point. This improved smoothness slightly but still frequently produced sheath that was either very much out of round or perhaps had flattened portions on the sides. This flattening was believed to be due to the easier distribution of pressure to the top and bottom rather than to the sides of the die and core-tube; thus a variable ex-

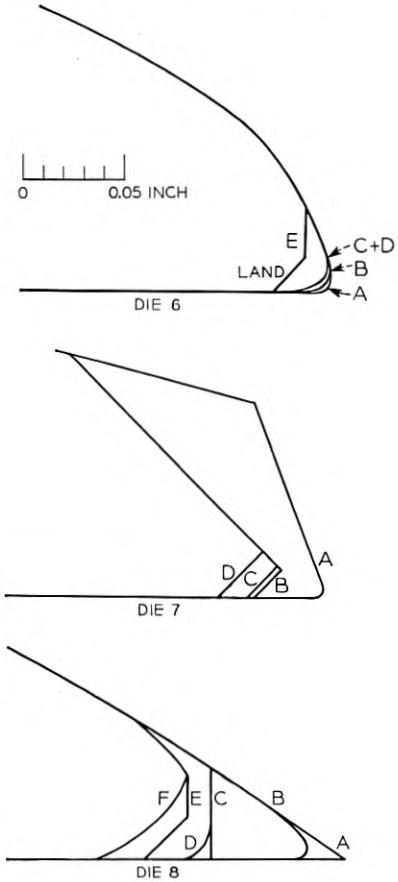


Fig. 7 — Longitudinal sections through upper tip of dies at various stages of modification.

trusion rate was produced around the annulus. In order to equalize this pressure the radius at the tip of the die was replaced with a 45° conical section ground as shown in 6E of Fig. 7. This provided a land whose length could be varied as required. A land, as the term is used in this paper, consists of a short conical ring section extending forward at some particular angle to the die axis toward the tip end of the die. This land, then, formed the outer sheath gate surface of the annulus, and produced a large improvement in both smoothness and roundness of the sheath when used in conjunction with core-tube 5D. However, an adverse effect on extrusion rate was noted. Since sharp shear edges had previously shown beneficial effects, another 45° cut was made as shown for 7B in Fig. 7. This change resulted in considerable improvement in extrusion speed. Fig. 10 shows typical data obtained when comparing the effect of two die shapes under conditions of constant pressure and core-tube design. In these two extrusions variations were made in both the sharpness of the leading edge of the die and the length of the restricting path

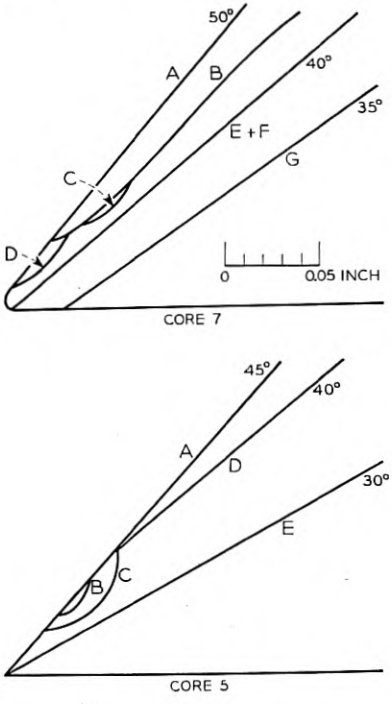


Fig. 8 — Longitudinal sections through upper tip of core-tubes at various stages of modification.

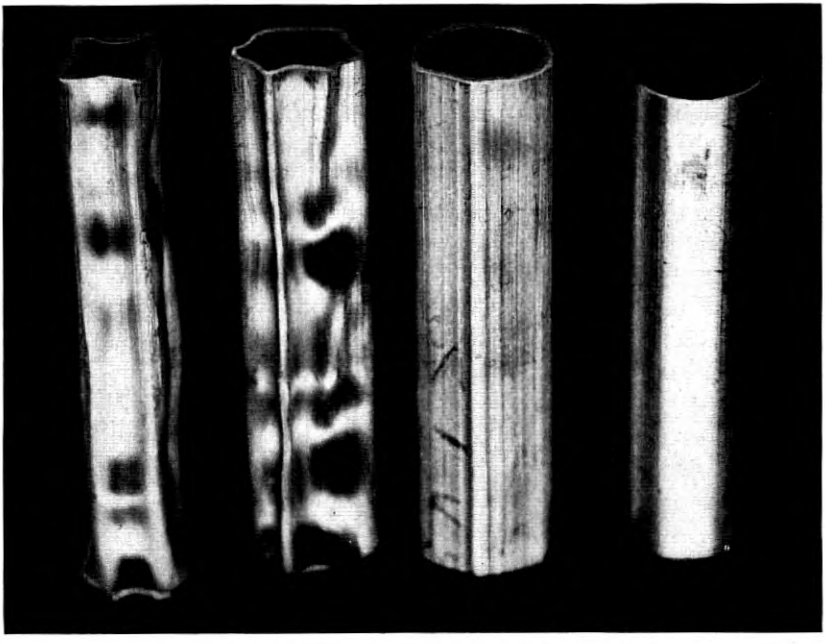


Fig. 9 — Evolution of tubing as extrusion tool contours were improved.

through which the aluminum must travel while being formed into sheath. Toward the end of the stroke the extrusion speeds through the two die shapes differ by a factor of about ten. While the extrusions start at constant temperature, the severe plastic working near the tips of the die and core-tube causes an increase in temperature. This is particularly true near the end of the charge where extrusion rates increase greatly. Exact instantaneous temperatures at this point are extremely difficult to measure. The abscissa designations along the top and bottom of the graph show the interrelation between specimen numbers and per cent of stroke.

Many other details of the die and core-tube design have been found to be of importance. Increasing the length of the land from 0.020 to 0.033 inches produces a reduction of extrusion speed. Decreasing the land from 0.020 inches below approximately 0.005 inches tends to give rough, irregular sheath. Increasing the angle of the core-tube beyond the 45° angle of the die land results in sheath that is rough and not round. The best results were obtained when the core-tube angle was slightly less than the die land angle which, in these studies, was fixed at 45°. This

produced a slight restricting effect as the aluminum approached the final point of wall thickness formation. The minimum angle is somewhat dependent on the strength of the steel and its resistance to collapsing under the extrusion pressure. Another effect of the angle of the core-tube is in establishing the exit direction of the formed aluminum. It is known that the outside diameter of the sheath is smaller than the minimum diameter of the die orifice as shown schematically in Fig. 11. The sheath is funnel-shaped to a distance of about $\frac{1}{8}$ to $\frac{1}{4}$ inch past the extrusion point. The exact curvature followed in this funnel-shaped region is a function of extrusion orifice contours, temperature, extrusion rate and purity of the aluminum. Thus, different grades of aluminum might require slightly different die and core-tube combinations.

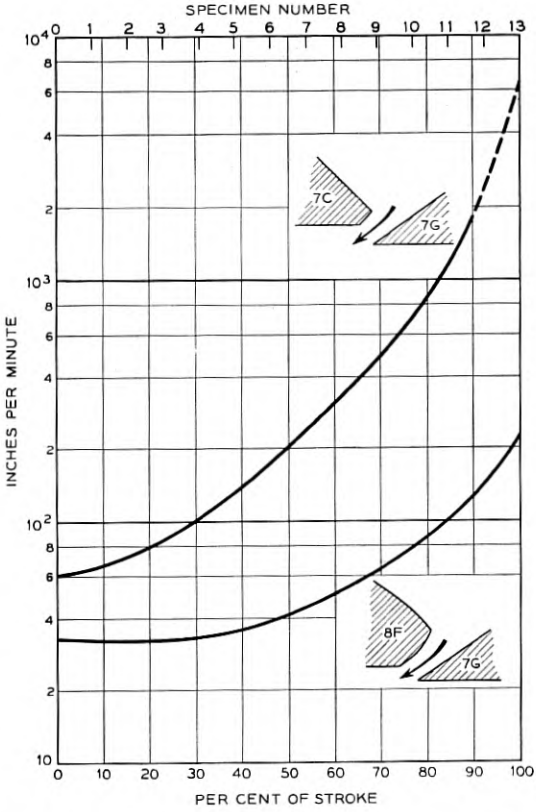


Fig. 10 — Effect of die shape on extrusion rate at 335°C and 62,500 psi. Note change of rate as extrusion progresses.

Another detail shown in Fig. 11 is the thin adherent layer of aluminum (cross-hatched sections) both on the land surface and a corresponding ring around the tip of the core-tube. The wetting of these surfaces appears to be of paramount importance for ease of extrusion and smooth-

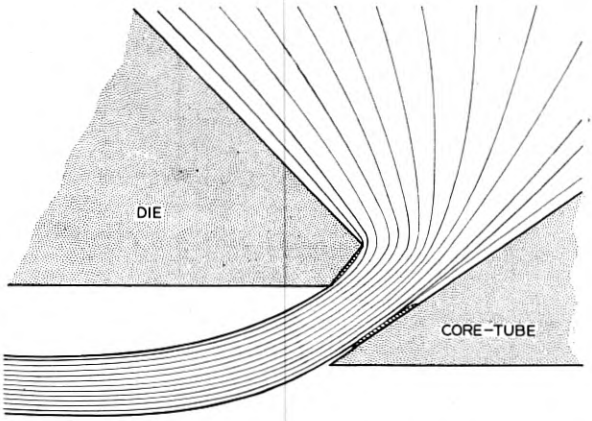


Fig. 11 — Schematic representation of flow of aluminum through a section of the extrusion orifice.

ness of sheath. The first few lengths produced by a clean coated die and core-tube are usually not as smooth as succeeding lengths. Until this annulus is wetted by aluminum, variations between slipping over the steel and shearing within the aluminum result in rough sheath.

EFFECT OF TEMPERATURE ON EXTRUSION RATE

Extrusion rates were obtained for several temperatures using die 7C and core-tube 7G and an applied pressure of 65,000 psi. These are plotted in Fig. 12 for a stroke position which is designated as specimen 2 corresponding to approximately 15 per cent of the total piston movement. A small increase in extrusion temperature results in a relatively large increase in extrusion rate. This emphasizes the need to operate at as high a temperature as is consistent with no detrimental effect on the paper insulation of the core.

EFFECT OF PRESSURE ON EXTRUSION RATE

Data on this subject are obtained by keeping the temperature fixed and extruding at different constant pressure levels. These data are treated as in the study of the effect of temperature. The curve obtained

is shown in Fig. 13, again for die 7C and core-tube 7G at 335°C. It would not be expected that this relationship would hold over an extended pressure range since below some limiting value, corresponding to the yield point of the material, practically no extrusion would take place. Once the conditions of reasonable extrusion rate are obtained much greater rates can be developed by relatively small increases in pressure. For example, a press that would produce sheath at 100 ft. per minute could, with only a relatively small increase in pressure, produce at 200 or 300 feet per minute.

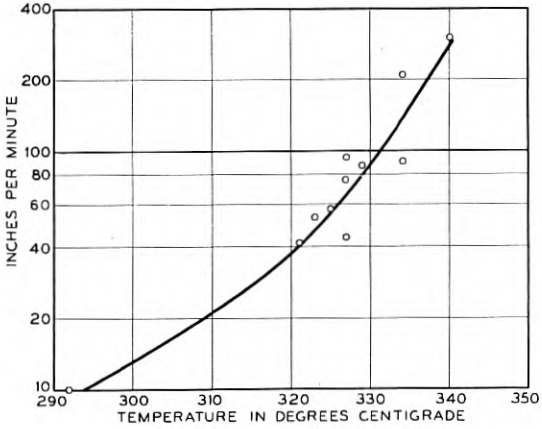


Fig. 12 — Effect of temperature on extrusion rate at constant pressure of 65,000 psi.

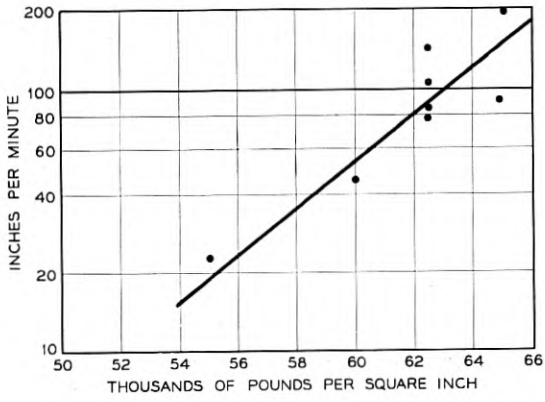


Fig. 13 — Effect of pressure on extrusion rate at 335°C.

EFFECT OF CHARGE LENGTH

The present studies, as well as reports of other investigators[8] of extrusion phenomena, show that considerable pressure is lost in forcing the billets to advance through the cylinders. Under some circumstances the material next to the wall slides when properly lubricated while in others the layer next to the wall remains essentially stationary and flow takes place by shearing within the aluminum. The degree to which these effects are present depends on the pressures required for extrusion. Under conditions of slight restrictive forces at the extrusion end of the cylinder, such as might be involved when a low extrusion ratio* is used, the pressure applied by the piston might be low. With these circumstances the pressure multiplied by the coefficient of friction would result in a frictional force less than the shear strength of the aluminum layer directly in contact with the cylinder wall. In these instances flow would take place by slip between the aluminum and the steel. When high extrusion ratios and correspondingly high pressures are used, the frictional force may exceed the shear strength of the aluminum, in which case flow takes place by shear within the aluminum. In either case there is a relatively rapid decrease in pressure at constant extrusion rate as the distance of the piston surface from the extrusion port becomes smaller.

The decrease of pressure with charge length is shown in Fig. 14 for several purities of aluminum at a constant extrusion speed of 25 feet per minute and a temperature of 331 $\pm$ 6°C. At this slow rate only a slight increase in temperature takes place at the die tip. The decrease in pressure is, therefore, nearly directly the result of decreasing charge length. The effects of purity are also apparent, of which more will be said later.

EFFECT OF TEMPERATURE ON EXTRUSION PRESSURE

This subject is of great importance since it affects the design and capacity of extrusion equipment. Data collected at a constant extrusion rate of 25 feet per minute are presented in Figs. 15 and 16 for 99.99 per cent and 99 per cent aluminum, respectively. The decrease in the slopes of the curves as the temperatures are raised or as the purity increases is evident. These slopes correlate most closely with initial pressure, which in turn is probably related to the shear strength of the materials and the consequent force necessary to push the aluminum billets through the cylinders. If the billets moved by sliding along the cylinder walls, the force would also be related to the initial pressure through the coefficient of friction. Since there is less loss in pressure due to shear and

* Cross-sectional area of the cylinders divided by that of the sheath.

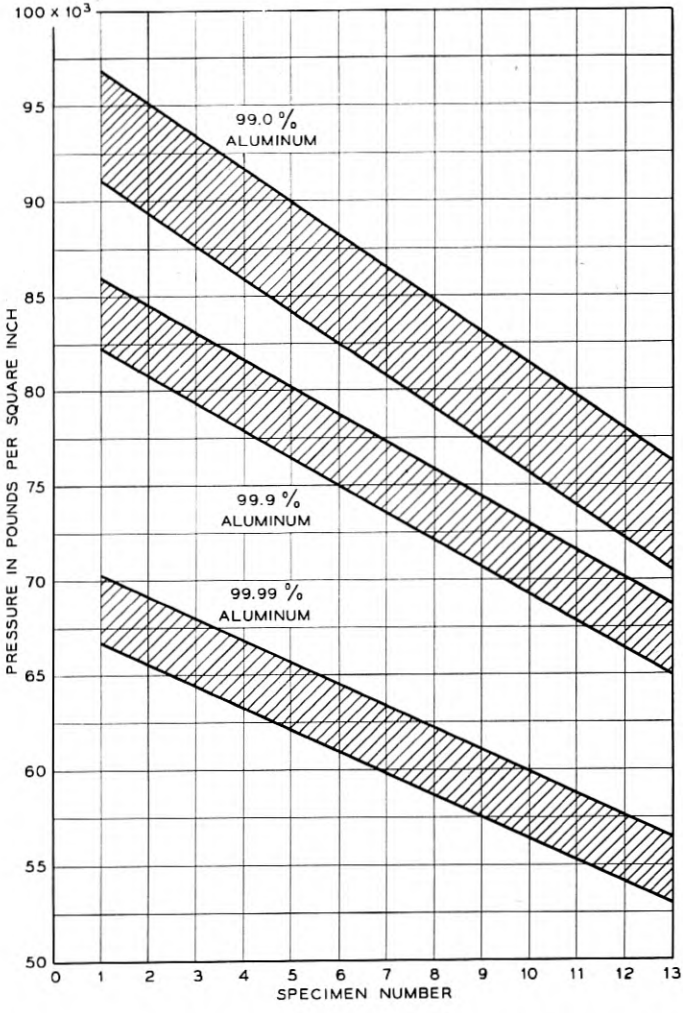


Fig. 14 — Effect of charge length and purity on extrusion pressure at constant rate.

friction at higher temperatures, there would thus be less need to maintain a low ratio of billet length to diameter. This is possibly the reason for some European success reported using long billets.[4] Fewer charge welds, press-marks and charging delays would result.

Further consideration of the pressure-temperature relationships gives some insight on the plastic behavior of aluminum during extrusion.

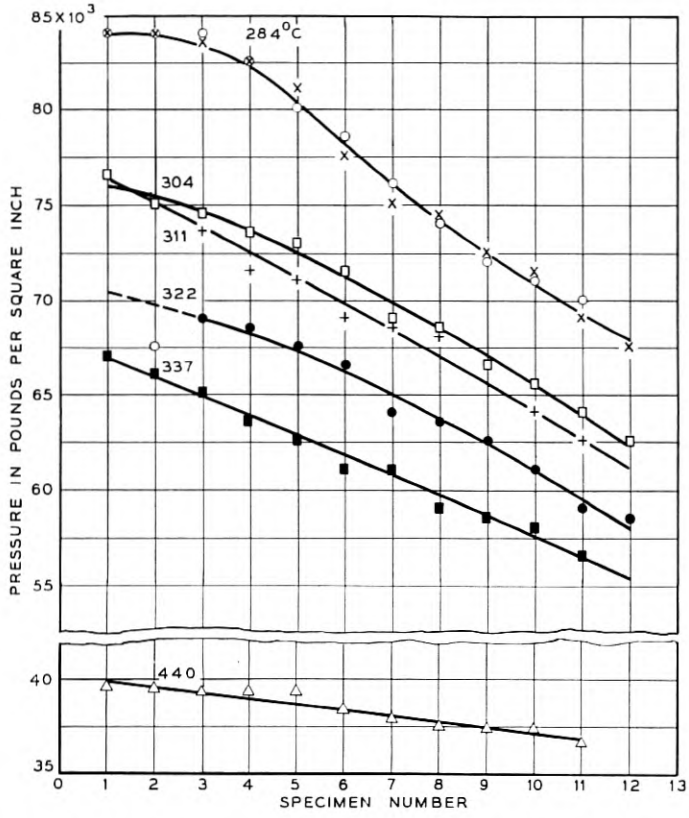


Fig. 15 — Effect of charge length and temperature on extrusion pressure at constant rate for 99.99 per cent aluminum.

Once the yield point of the aluminum has been well exceeded, it might be expected that it would flow in a manner analogous to true plastic behavior. A plasticity relationship would normally follow an equation of the Arrhenius type:

$$\text{Log } \eta' = \text{Log } K + \frac{E}{2.3RT}$$

where η' is the plasticity, K is a proportionality factor, and E is the activation energy for plastic flow. If the aluminum were behaving in this manner, the pressure necessary to force a definite volume of aluminum through a specified orifice in unit time would be directly proportional to the plasticity. After about the fifth specimen has been extruded,

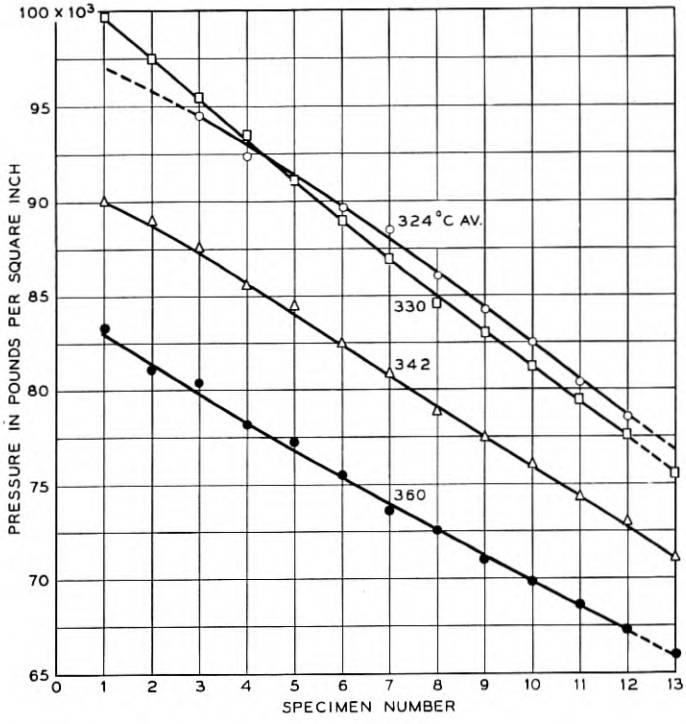


Fig. 16 — Effect of charge length and temperature on extrusion pressure at constant rate for 99 per cent aluminum.

relatively stable conditions are established and the data may be compared further. When the logarithms of pressures for this specimen are plotted against the reciprocal of their absolute temperatures, the straight line shown in Fig. 17 is obtained for 99.99 per cent aluminum. The data for the 99 per cent are not sufficiently extensive to fix the exact slope but it is probably similar. This linearity indicates that aluminum has a plastic behavior under the conditions of temperature and pressure used in these experiments. Other indications point to a tendency to form a stationary layer at the cylinder wall and die parts similar to viscous behavior. Assuming such a relationship, the activation energy for plastic flow may be calculated as 3.6 kilo-calories. This is a low value compared to activation energies for viscous flow of $\sim$ 10 kilocalories for a soft rubbery plastic such as polyisobutylene at 217°C, or 150 kilocalories for glass at 600° to 700°C.

PURITY OF ALUMINUM

Several factors enter into the choice of grade of aluminum for cable sheath. Among them are ease of extrusion, cost, availability, corrosion resistance and mechanical properties. The extrusion behavior has been shown in Figs. 14 through 17 and is summarized in Fig. 18. The first increments in impurity have proportionately much greater effects on extrusion pressure than subsequent increases have. Our plasticity studies indicate that iron is more responsible for increase in pressure than are the copper and silicon which are also normally present in commercial aluminum. The total spread of pressure required over the purity range studied is enough to influence the design of extrusion equipment. The maximum pressure for even the least pure material is not beyond commercial feasibility.

There is a large differential in cost between 99.99 per cent and 99.9 per cent aluminum. The difference between 99.9 per cent and 99 per cent is much less. The same relationship holds true for availability. There is at present no major source of the super-pure material in this country.

Tensile tests have been made on extruded pipe produced from the three grades of aluminum studied. The results are noted in Table I and shown graphically in Fig. 19 for specimens extruded at about 330°C.

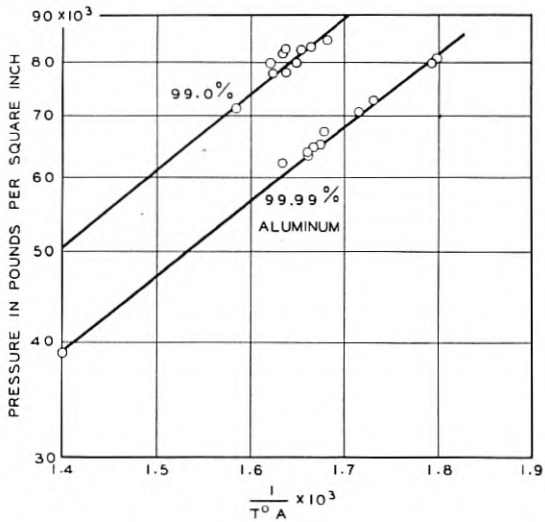


Fig. 17 — Dependence of log pressure on reciprocal of absolute temperature.

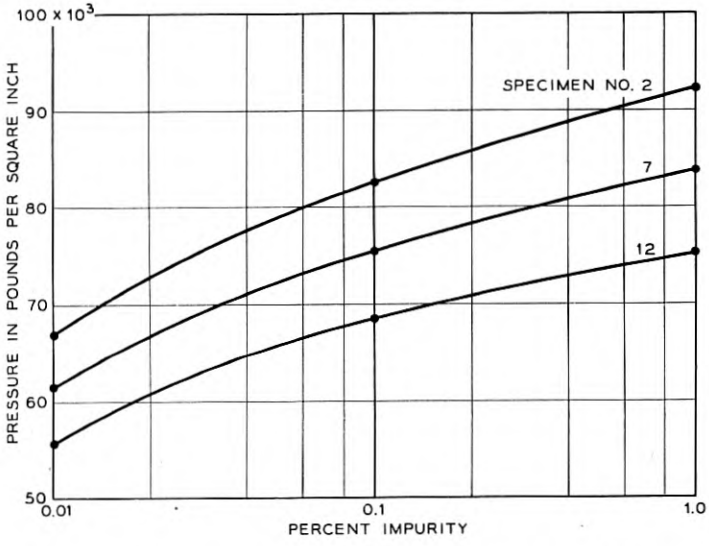


Fig. 18 — Effect of impurity on extrusion pressure at constant speed.

The tests were made by gripping flattened ends of pipe in the jaws of a testing machine. The gage length used was 10 inches and the crosshead was moved at the rate of 1 inch per minute. Stress-strain curves were plotted automatically by the machine.

TABLE I — TENSILE STRENGTH OF THREE GRADES OF ALUMINUM

	Tensile Strength (psi)			Elongation (per cent in 10 inches)		
Purity in Per Cent	99.99	99.9	99.0	99.99	99.9	99.0
Specimen No. (Position in Charge)						
1	9,750			45.5		
2	10,000	11,600	17,400	45.0	21.7	24.2
3	10,300	11,120	16,500	44.3	26.2	24.2
4	10,420	12,200	16,400	42.9	29.0	28.0
5	10,540	12,120	16,300	43.9	30.0	27.0
6	10,700	12,200	16,000	42.5	33.5	27.0
7	10,850	12,200	15,900	42.9	36.5	
8	10,750	12,080	15,900	42.5	27.8	
9	10,630	12,200	15,600	42.6	34.3	
10	10,500	12,150	15,600	40.4	24.0	
11	10,200		15,450	42.8		
12	10,000		15,250	39.7		
13			15,100			

The increase in tensile strength and the decrease in elongation with increasing impurity are of the magnitude to be expected. The variation of strength along the length of the charge may be related to differences in degree of recrystallization observed in the microstructure. In the case of both the 99.9 per cent and the 99 per cent materials the ratio of amounts of impurity in and out of solution may also have an effect on the strength.

Typical stress-strain curves for the three materials are reproduced in Fig. 20. The abrupt change in direction in the early part of the curve for the highest purity material represents a yielding of the coarse soft grains while work hardening capacity is being built up. The less pure materials have a finer grain structure, and, having a higher recrystallization temperature, may well retain some of the strain resulting from extrusion.

EFFECTS OF EXTRUSION TEMPERATURE ON CORE WRAP PAPER

A major consideration in the production of aluminum cable sheath has been the need to limit the extrusion temperature in order to prevent damage to the core wrap paper. An extrusion temperature of 325°C was found to result in slight discoloration of red and white undried core wrap paper. At this temperature, when the press is stopped in the order of 30 seconds for recharging, the paper assumes a distinct brown coloration but does not appear to be embrittled. Studies were made of the

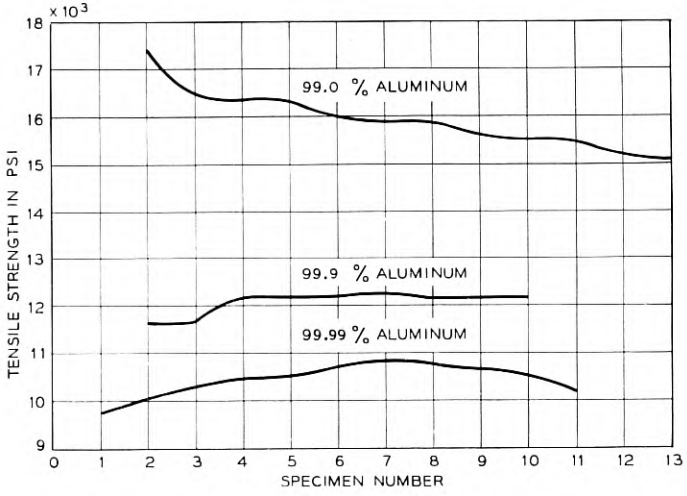


Fig. 19 — Tensile strength of extruded aluminum pipe.

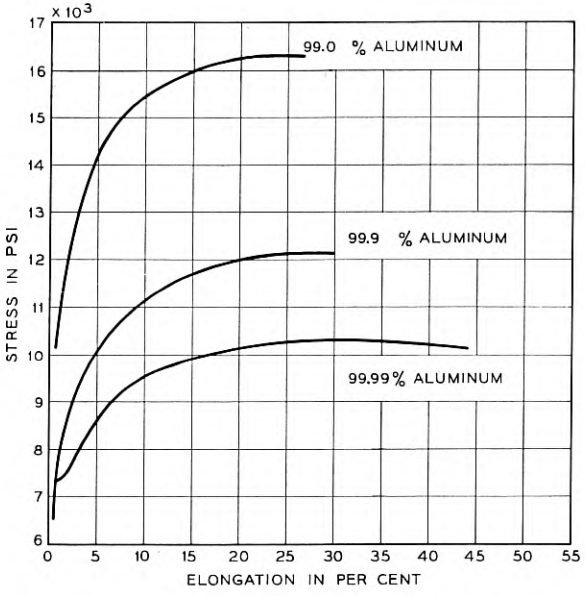


Fig. 20 — Stress-strain curves for extruded aluminum pipe.

temperatures existing in the core by attaching thermocouples to the outside, under one, and two layers of paper. The thermocouples were connected to sensitive recorders and showed the following results when an extrusion temperature of 325°C was used:

Couple on outside of core: 279 to 295°C.
Couple under one layer of paper: 225 to 235°C.
Couple under two layers of paper: 197 to 215°C.

JUNCTIONS BETWEEN CHARGES

At the outset of the laboratory studies it was realized that considerable trouble might be expected when successive billets were added to cylinder residues. In ordinary commercial extrusion where continuous lengths of tubing are not required, the common practice is to remove the cylinder residue from the press after each extrusion stroke. Attempts to introduce successive billets have often resulted in poor welds or pinholes in the tubing. Beside contamination from improperly prepared surfaces, another basic source of junction difficulty may occur. When a cylinder of plastic material such as aluminum is compressed between parallel plates, "barreling" takes place due to the restrictive effect of friction on the two ends of the cylinder. Thus, when a cylindrical billet

is placed in the press cylinder, there is a tendency for the billet to become thicker in the middle as pressure is applied to one end. This may effectively seal the cylinder at this point and entrap a fairly large amount of air between this restriction and the old charge material. It was found that this situation could be corrected to a major degree by scoring the billet with four equally spaced longitudinal "V" shape grooves approximately 0.010 inches in width and depth. When pressure is first applied to a billet so scored, "barreling" takes place as before, but the pressure against the cylinder wall at the mid-point of the billet is not sufficient to seal off the inscribed grooves. Continued application of pressure to the piston then forces all portions of the billet into contact with the previous charge and the cylinder walls. Final increase in pressure to near extrusion value causes the grooves to close thus trapping a minimum amount of air. Sheath and pipe produced under this technique have been tested under 60 psi air pressure and found to be free from pinholes.

Air entrapment can be further minimized by maintaining the preheating temperature of the billets slightly below that of the residual charge in the press cylinder. The end of the billet in contact with the hotter residual charge increases more in diameter than the cooler portions closer to the piston, thus reaching the cylinder walls near the old charge residue first. As temperature equilibration takes place under pressure, the increased diameter moves outward along the axis of the billet toward the piston forcing the air out of the cylinder. A combina-

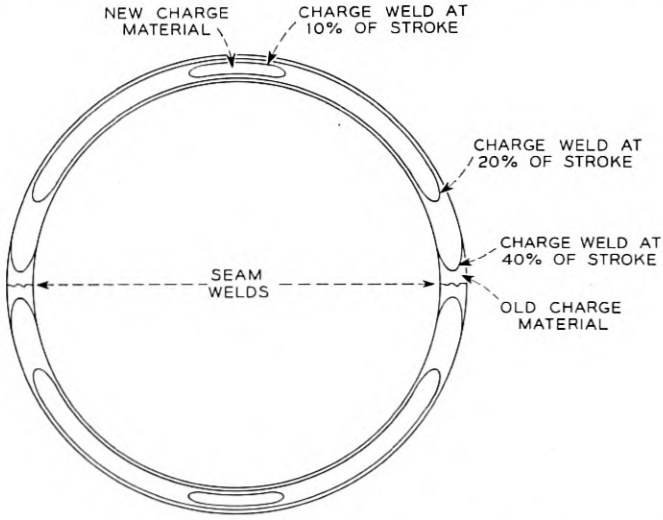


Fig. 21 — Diagram of sheath cross-section at various stages of extrusion.

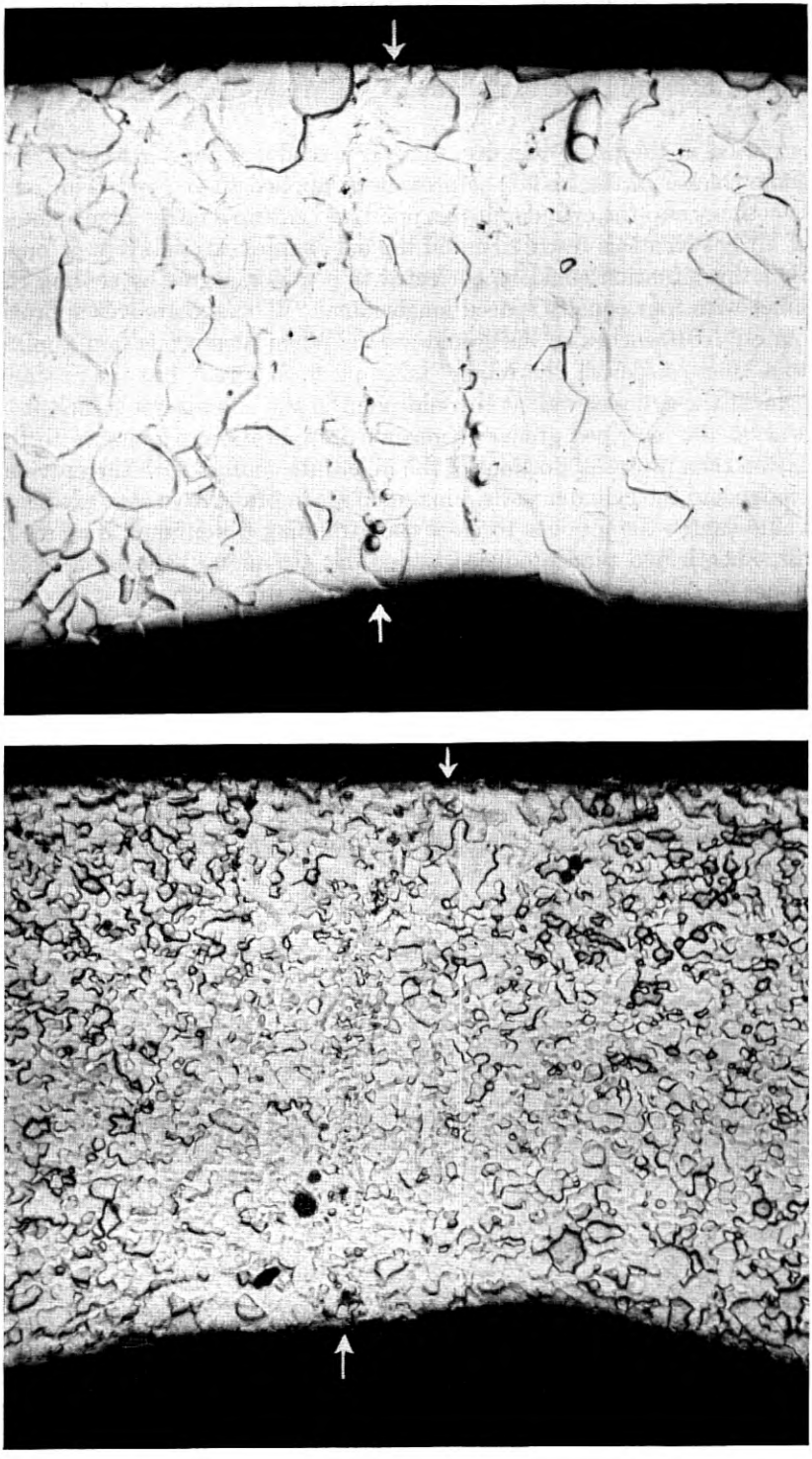


Fig. 22 — Transverse microsection of extruded 99.99 per cent aluminum sheath. Above, seam weld coarse grained material, and below, seam weld fine grained material, both at 150x.

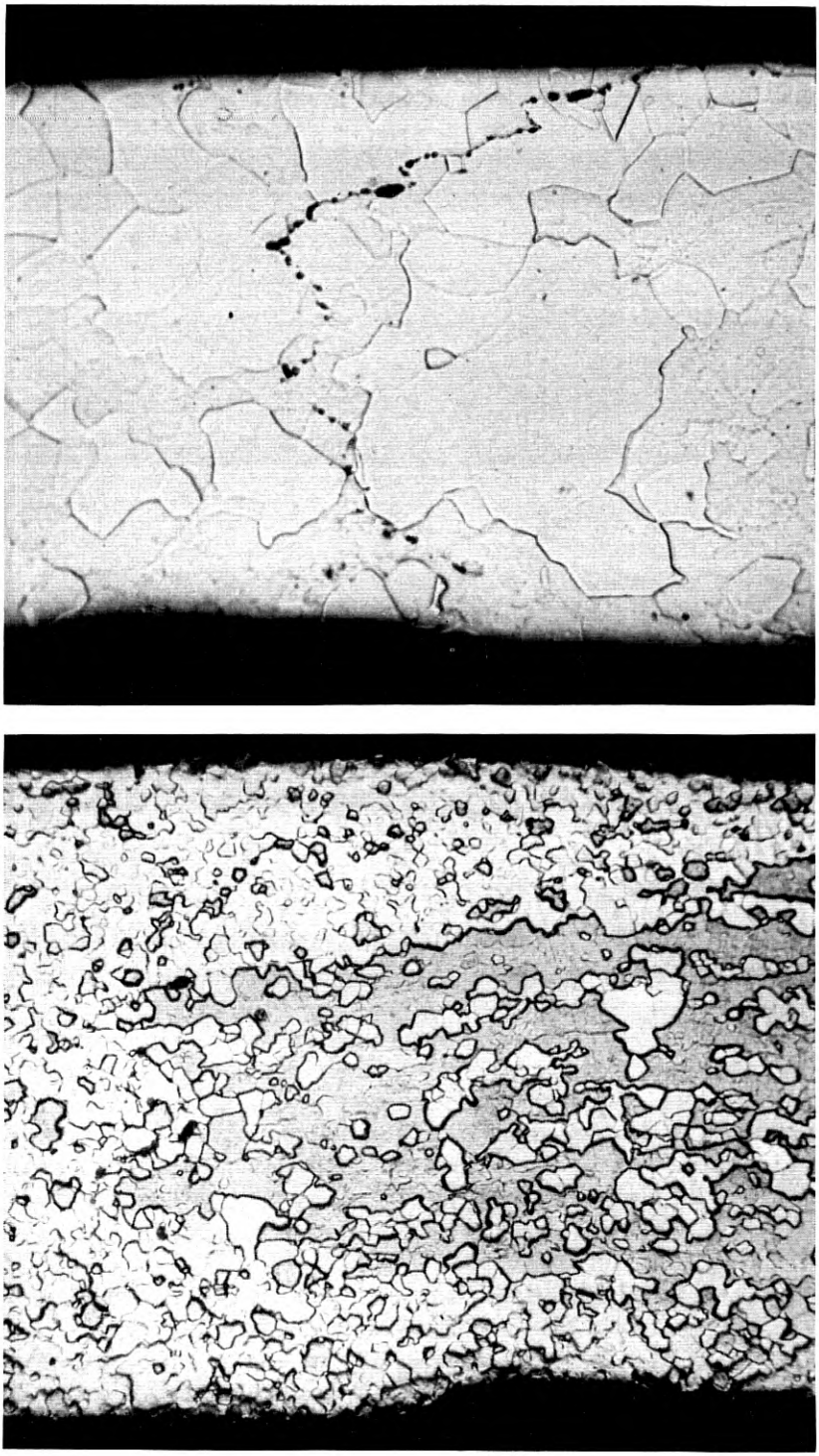


Fig. 23 — Transverse microsection of extruded aluminum sheath. Above, charge weld — line of demarcation is plainly visible. Below, charge weld — the new charge material appears darker because the grains have a different orientation. Both at 150x.

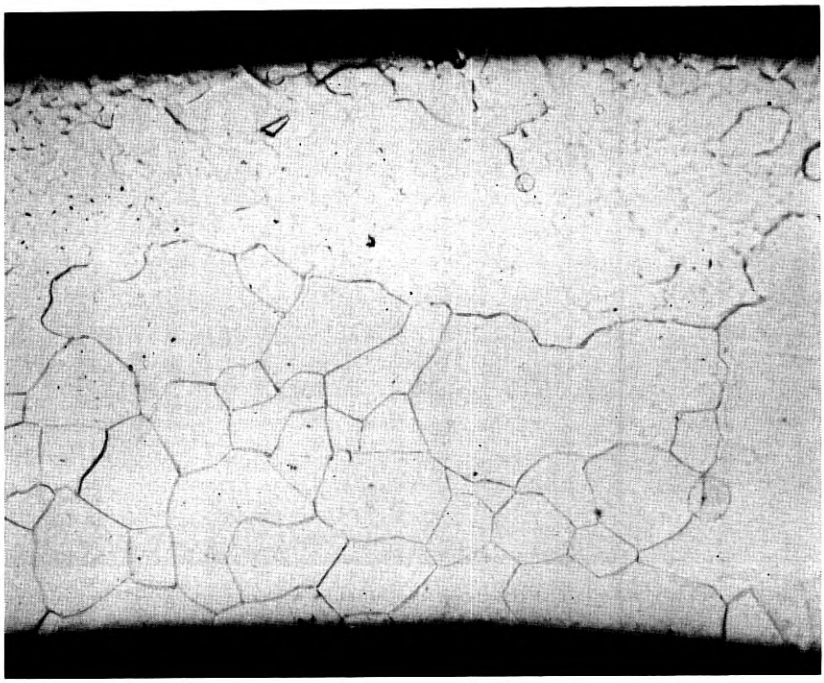


Fig. 24(a) — Top of sheath away from welds. Material near top of photo is not recrystallized. At 150x.

tion of these techniques coupled with care in preparing billets and elimination of lubricant during the charging operation has given tubing free from holes.

Samples cut from various positions along the length of pipe extruded from a single charge have been subjected to expansion tests using a tapered steel plug similar to the technique used in testing welds in lead cable sheath. The welds have proved sound and expansions of 30 to 40 per cent are found. The location of the breaks occurred at random with respect to the position of the welds.

MICROSCOPIC EXAMINATION

As would be expected from the double ingot charging procedure, seam welds are present in the aluminum sheath diametrically opposite on a horizontal plane as shown schematically in Fig. 21. The progressive displacement of the old charge residue is also indicated. If considerable dross is present, there may be sharp lines of demarcation. Fig. 22 shows

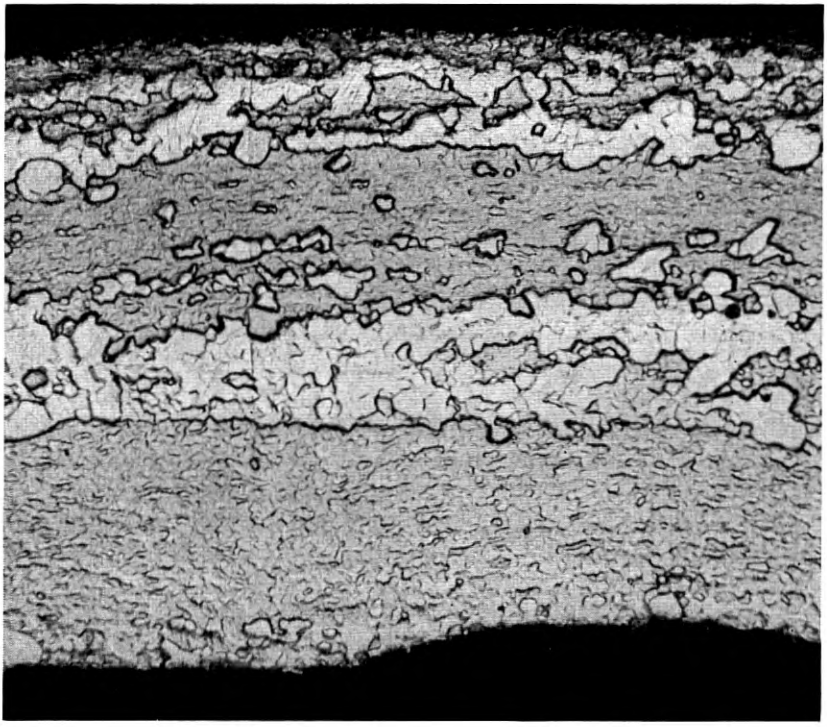


Fig. 24(b) — Away from weld region. Partial recrystallization, 150x.

average conditions at the seam welds. The interlocking of the grains across the junction indicates good welding. The samples were prepared by electropolishing followed by anodizing to show the grain structure. The difference in grain size is a function of the extrusion conditions. Fig. 23 shows charge welds after about 25 per cent of the charge has been extruded. In the upper picture a distinct line of oxides and voids is present. The weld, however, is still quite sound as was shown by expansion tests. In the lower photograph practically no inclusions are present. The junction is visible because of the difference in orientation of the new charge material which causes it to appear darker. Two sections away from the weld region near the top of the sheath are shown in Fig. 24. Here again the marked difference in grain size is visible. In addition, differences in grain structure are apparent in each sample. The crystal structure in the lower half of Fig. 24(a) which corresponds to the inside of the sheath is well defined. The indistinct structure in the

outer region is caused by lack of recrystallization after extrusion. Examination under polarized light shows preferred orientation and lineage in this region. The irregularity in recrystallization is even more pronounced in Fig. 24(b). The extrusion temperature range used is very near to the recrystallization temperature for material of this purity. The extent of recrystallization is also a function of the amount of plastic deformation prior to exposure of the material to maximum temperature as it passes through the die orifice. Since the cross-section may be composed of material from the center and outer portions in different proportions which would undergo considerable variation in deformation before reaching the point of sheath formation, the structural differences are not surprising. Structure varies along the length of charges due to the combined effect of deformation variation and the temperature cycle usually accompanying an extrusion cycle especially when fast extrusion rates are used. There are also some indications that the structure of the charged billets has an effect on the extruded pipe. The photomicrographs shown are for super purity aluminum. As would be expected, grain size and degree of recrystallization decreases with decreasing purity for a given set of extrusion conditions.

Apart from low pressure advantages gained by extrusion at temperatures above 330°C, more uniform grain structure might be obtained. Low temperature extrusions often result in mixtures of coarse and fine grained material. The latter may be either imperfectly recrystallized and filled with much lineage or may have the badly distorted appearance of unrecrystallized material. It is not known if this mixed structure might have an adverse effect on service life. It is reasonable to suppose, however, that some benefits might result from uniformity in grain size.

PREPARATION OF BILLETS

The high purity aluminum was received in two different forms; either direct chill castings, three inches in diameter and approximately four feet long or pigs which were melted at the Laboratories and cast into graphite molds. The 99.9 per cent aluminum also was melted and cast. The 99 per cent aluminum was in the form of extruded rod.

All of the cast surfaces, even those produced by direct chill, showed some folds and ripples. Those cast in graphite molds had in addition a certain amount of oxide inclusions and voids. In order to minimize inclusions in the final billet, the cast surfaces were removed by shaving. The ingots were first sawed to suitable lengths, then heated to 400°C. At this temperature, they were placed in a hydraulic press and a sharp

edged shaving tube forced over them lengthwise. This removed about ¼ to ½ inch of material from the outer surfaces. The resulting billets were generally smooth and clean after this operation. The surface was far superior to those observed on billets to be extruded in commercial plants even though some of their billets had been turned in a lathe. Before the billets were used, the surfaces were cleaned with trichloroethylene to remove traces of lubricant left from the shaving operation. Generally, the billets were stored in an oven at 300°C prior to insertion into the cylinder of the extrusion. In conformance with commercial practice in handling aluminum billets, no attempts were made to prevent oxidation during any of the melting or subsequent heating operations.

The start, middle, and end of a shaving operation is shown in Fig. 25. Fig. 26 shows the shaving tool partly removed from the billet at the end of an operation. The force required for shaving is about 20 tons. It seems probable that such a procedure would lend itself to commercial operation. Chip breaking knives and cutters to form the longitudinal grooves on the billet would be desirable on a commercial shaving tool.

WEAR ON PRESS PARTS

It was anticipated that the aluminum would cause a certain amount of erosion on the steel parts of the press. However, during experimental operation this has not been evident. For example, the clearance between the piston and the cylinder wall originally was approximately 0.002 inches and through this space a slight back extrusion has always taken place. During many months of press operation, the amount of this back

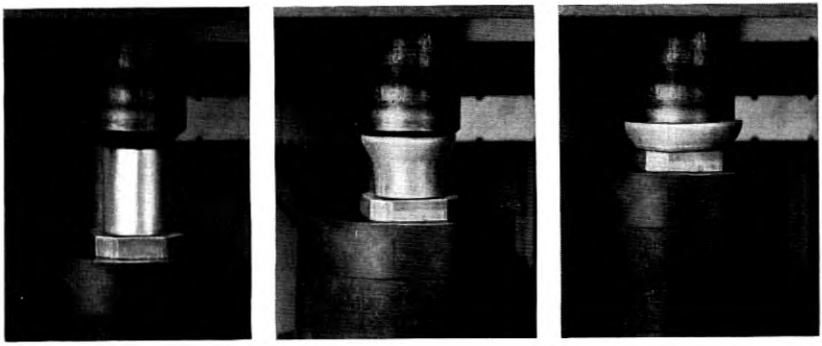


Fig. 25 — Start, middle, and end of shaving operation for removing outer surface of cast aluminum billets.

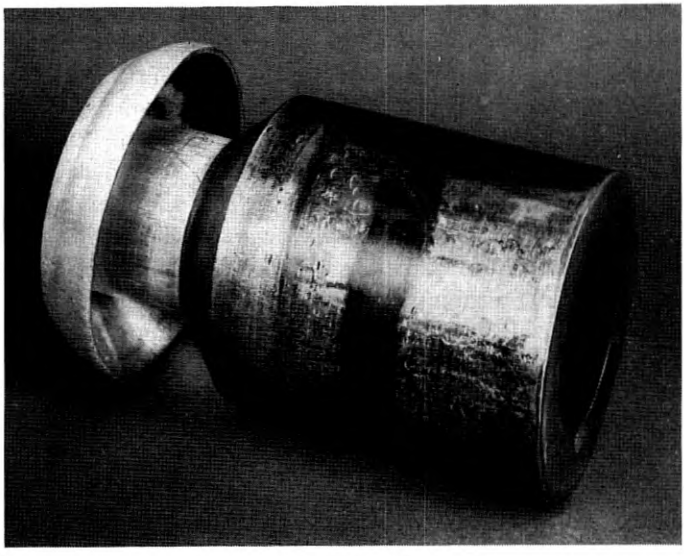


FIG. 26 — Shaving tool partly removed from billet after operation.

extrusion or scarf has not increased visibly, indicating no wear on cylinder or piston edge which would have been very noticeable here. Examination of the core and die surfaces also have shown little, if any, erosion and there has been no evidence of alloying between aluminum and steel.

Attack and wear by the aluminum is probably prevented by the relatively stationary layer of aluminum which adheres to the steel surfaces during extrusion. The large proportion of motion takes place by shear within the aluminum rather than by sliding over the steel surface.

LUBRICATION

The initial experiments showed that application of oil coatings to successive billets had a favorable effect on decreasing the extrusion pressure. However, this practice invariably resulted in poor welds in the sheath and was discontinued. Over a period of many months no lubricant of any type was applied to the billets which were used for recharging the press. Coatings on the core-tube and die, however, have beneficial effects on press operation. Among the materials investigated as die and core-tube lubricants were Teflon coatings baked on at 350°C, various silicone greases, mixtures of petroleum jelly and mutton tallow, heavy paraffin-base cylinder oil, copper flashes produced by replacement from especially prepared copper bearing solutions, and finally a heat polym-

erized oil coating. This latter coating, which proved to be by far the most effective, was produced by dipping the parts in high temperature motor oil and then heating them in air to an estimated 350°C by a radiant heater. The resulting partly-polymerized, partly-carbonized hydrocarbon produced a very adherent, hard, black, glossy finish which has proved of considerable value in reducing adhesion of aluminum to the extrusion surfaces. While some of the coating is removed as a result of continued extrusion, the parts so coated are relatively free from massive aluminum adhesions when removed from the press. The coating did not prevent the beneficial wetting by aluminum of the land and opposite core-tube region (Fig. 11).

SIZE FACTOR

Because of the complexity of pressure distribution which causes the flow through the extrusion passages, it is difficult to calculate the forces to be expected in different shapes and sizes of extrusion equipment. This anomalous pressure distribution was demonstrated by the existence of voids at particular locations in the cylinder residue after considerable extrusion of lead at room temperature under a pressure of 20,000 psi. Considerable data[8] and theoretical considerations point to a straight line relationship between the logarithm of extrusion ratio and the extrusion pressure as shown in Fig. 27. For this reason, if attempts were made to

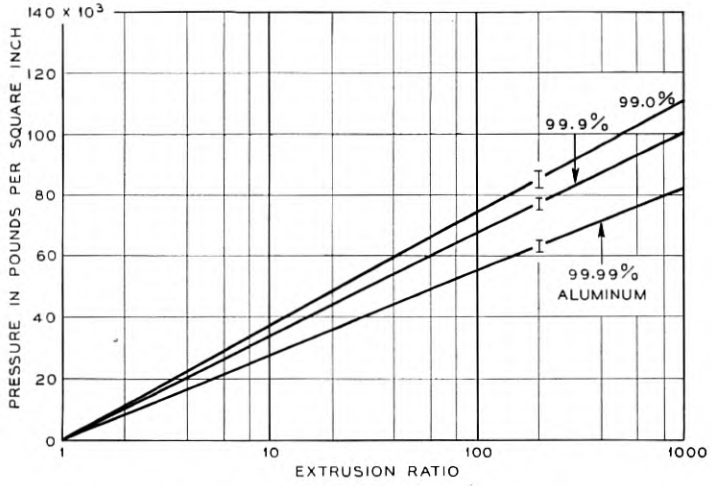


Fig. 27 — Effect of degree of reduction on extrusion pressure at constant speed.

extrude ½ inch diameter sheath using a press having a cylinder five times the area of the experimental laboratory press, it could be estimated that a pressure increase from 63,000 psi for the laboratory press to 82,000 for the larger press could be expected for 99.99 per cent aluminum. For 99.9 per cent aluminum, the increase which might be expected would be from 77,000 to 101,000 psi. With 99 per cent aluminum the increase would be from 85,000 to 111,000 psi. These values are all well below the 150,000 psi figure commonly used in commercial extrusion. Were some of the larger sizes of sheath to be produced, the extrusion ratio would decrease and result in a correspondingly lower extrusion pressure. Up to the present time, sheath of only one size has been extruded on the Laboratories press. It has, therefore, not been possible to confirm the extrusion ratio-pressure relationships on the two piston type of press. It might be emphasized in connection with Fig. 27 that no account has been taken of the 90° change in direction so that the extrusion pressure at a 1 to 1 ratio would most certainly be somewhat higher than zero pressure. An increase at this point in the abscissae would pivot all curves about the intercepts at 200 and lower the extrapolated points at higher ratios.

SUMMARY

A survey has been made of the factors involved in extruding aluminum directly over paper insulated cable core. Since the upper temperature limit is restricted by the presence of the paper and the force to deform aluminum is many times greater than that for lead, conventional lead sheathing equipment cannot be used. An experimental two-opposed piston arrangement with die and core-tube located on an axis 90° to that of the pistons and midway between them has been found to operate at much lower pressures. The shapes of die and core-tubes have been shown to be critical with respect to pressure required and perfection of the sheath. Data are presented to show the effect of purity of aluminum on the interdependence of temperature, pressure and extrusion rate. Special lubrication coatings for the die and core-tube are discussed. Entrapment of air between successive charges was prevented by placing longitudinal grooves on the billets. These were then charged at a temperature somewhat lower than that of the press. Using these techniques, the seam welds produced were as strong as the rest of the sheath. Tensile properties of some of the extruded sheath are given. Microstructures of extruded sheath are shown.

ACKNOWLEDGEMENT

The authors wish to express their appreciation for the encouragement and advice of E. E. Schumacher, to R. E. Elicker who was associated with the early studies, and to M. J. O'Brien who aided materially in obtaining the data.

REFERENCES

1. G. Sachs and W. Linicus, Pressure Requirement and Flow Processes in Extrusion, Mitteilungen der deutschen Materialprüfungsanstalten — Special Issue XVI, pp. 67–96, 1931.
2. W. Deisinger, Production of Aluminum Cable Sheath, Zeit. f. Metallkunde **31,** pp. 305–310, 1939.
3. K. S. Wyatt, Sheathing of Cables With Aluminum, Wire and Wire Products, Oct., 1953.
4. H. Borchardt and G. Hosse, Cable With Extruded Aluminum, Siemens-Zeitschrift **4,** pp. 203–211, 1953.
5. Aluminum Cable Sheathing, Electrical Journal, pp. 1951–1952, May, 1954.
6. Cable Sheathing by Direct Extrusion, Light Metals, pp. 214–215, July, 1954.
7. G. M. Bouton and G. S. Phipps, Compression of Lead Alloys at Extrusion Temperatures, Proc. A.S.T.M., **51,** pp. 761–770, 1951.
8. C. E. Pearson, The Extrusion of Metals, John Wiley & Sons, N. Y., 1949.

A Tunable, Low-Voltage Reflex Klystron for Operation in the 50 to 60-kmc Band

By E. D. REED

(Manuscript received January 19, 1955)

Electrical and mechanical techniques are described which have been successfully applied to the design of a reflex klystron operating in the 50,000 to 60,000-mc band. Not only have these techniques resulted in a practical tube operating at the highest frequency yet achieved with a gridded, low voltage reflex klystron, but they point the way to future work at still higher frequencies.

The M1805 reflex klystron has a mechanical tuning range of over 10,000 mc and at a beam voltage of 600 volts delivers a maximum CW power of 15–30 milliwatts. Its electron-optical system resulted in a repeller mode exhibiting good symmetry and almost complete absence of electronic hystersis.

CONTENTS

	Page
Introduction	563
General Design Considerations	566
The Electron Gun	568
The Passive Circuit	575
Grids	584
The Reflector	589
The Electrically Significant Contours	590
Considerations in Choice of Tuner	591
The Mechanical Structure	593
M1805, Electrical Performance	595

INTRODUCTION

The design of a low-voltage tunable reflex klystron for continuous operation in the millimeter waveband poses a number of difficult problems most of which stem either directly or indirectly from the small physical size of the resonant cavity. In addition, we must contend with the unpleasant yet inescapable facts that klystron cavities become inherently poorer with increasing frequency and that both smoothness and freedom from stresses in cavity walls assume increasing importance.

Finally, if a long-life, oxide coated cathode is desired, that is, one operating at moderate surface current density, a somewhat delicate gun design problem must then be faced.

Many techniques, both electrical and mechanical, successfully used in the centimeter region fail when applied to the design of a millimeter-wave reflex klystron so that we are forced to look for alternatives, i.e., specialized solutions which satisfy the requirements peculiar to this frequency range. Also, since many of the problems are mechanical in nature, it is obvious that the electrical design of a tube of this type cannot proceed without unusually close attention to structural feasibility. Another obstacle to tube development in a relatively new frequency band is the lack of broadband waveguide components normally available for the determination of important tube parameters.

The M1805* reflex klystron is the outgrowth of development effort which was aimed at exploring the possibility of millimeter-wave generation by means of a low voltage tunable reflex klystron. The ease of operation and general convenience afforded by such a tube not only were needed for the millimeter wave research in progress at Bell Laboratories and in the military establishment but, it was felt, would render practical and stimulate new investigations held in abeyance because of the lack of suitable primary signal generators.

As described in a recent paper by S. E. Miller,[1] the Bell System's interest in the millimeter wave region has been centered chiefly around the use of circular waveguide as a low-loss communication medium. This interest and, equally important, certain military needs led to an intensification of the millimeter klystron effort, an effort which, since the later part of the war, had been pursued at Bell Laboratories on a comparatively small scale and with several interruptions. As early as 1945, J. R. Pierce[2] succeeded in producing several milliwatts of continuous millimeter wave power. His tube, the 1464XQ, could be thermally tuned from about 45 to 48 kmc and with a resonator voltage of 400 volts, a value, which was determined by the power handling ability of the grids, would deliver some 2–5 milliwatts of RF power. A small number of these tubes was built but development was stopped short of completion. The experience gained in the course of that investigation and in later work of C. T. Goddard, however, has in no small way contributed to the success of the M1805.

The present development program, which was culminated in the highest frequency reflex klystron of the low voltage, gridded type ever

* This tube was developed under Office of Naval Research Contract Nonr 687(00).

built, started out with two distinct aims. The first was to determine whether a low-voltage, tunable reflex klystron for CW operation in the 5–6 mm band could be built on a laboratory basis. There was no assurance that this could be done within the limitations imposed by such requirements as low voltage operation, tunability, moderate emission density and long life. Moreover, as the performance of Pierce's 1464XQ had been marginal, scaling to a still higher frequency appeared a somewhat risky endeavor. The second objective was contingent upon the success of the first. It called for a tube of reproducible characteristics and one which would lend itself to quantity production if and when required.

Results obtained to date with representative numbers of M1805's indicate that the original design objectives have in most instances been met and in some, such as power output and tuning range, exceeded. A typical tube will tune mechanically from 50 to 60 kmc, will deliver a maximum of 15 to 30 milliwatts of RF power* within this band and will exhibit a clean and symmetrical mode shape, almost free from electronic hysteresis. Optimum coupling to a matched waveguide has been achieved over most of the mechanical tuning range and, in general, the tube seems to be as well behaved and as easy to operate as lower frequency reflex klystrons of more conventional design.

In its present form, the M1805 reflex klystron combines results obtained in the course of a number of separate studies all of which had to be essentially completed before the final mechanical design of this tube could even be started. These studies pertained to the electron-optical system and, more particularly, to the design of a highly convergent electron gun; they pertained to the evolution of an efficient passive circuit consisting of the tunable resonant cavity and the broad-band output coupling system. Finally, they were directed at finding methods of handling the tolerances and the dimensions — some of which resemble "normal" tolerances — dictated by the operating frequency of this tube.

As a result of these studies we found that the minute dimensions and tight tolerances encountered in a millimeter wave klystron do not present undue difficulties if handled with appropriate techniques and we have, in fact, obtained a degree of reproducibility in such parameters as gun perveance and transmission efficiency considerably higher than has been customary with lower frequency tubes. The consistency of the RF performance has been satisfactory too, although capable of further improvement.

Of the mechanical techniques responsible for this unusual measure of

* This spread includes a possible uncertainty in power measurement of about 1 db.

reproducibility and control in the face of minuteness both in the size of elements and their spacings, the more important ones are: (1) the extensive use of precision hubbing, (2) the reliance on optical rather than self-alignment to obtain the close degree of concentricity required, and (3) the use of individually selected ceramic spacers to determine spacings. Equally important, of course, has been an electrical design which permits the fullest utilization of these techniques.

GENERAL DESIGN CONSIDERATIONS

Suppose we now take a somewhat more detailed look at the major problems which confront the designer of a millimeter wave reflex klystron. Later in this paper we shall see what solutions were obtained to these problems and how they were applied to the design of the M1805.

Let us first examine the frequency determining resonant cavity. Its linear dimensions are directly proportional to wavelength or inversely proportional to frequency. Hence, a 4,000-mc cavity, for example, having an outer diameter of about one inch would shrink to one having a diameter of $\frac{1}{15}$ inch or 67 mils* if we wanted to go up in frequency by a factor of 15, that is, to 60,000 mc. Whereas a dimensional tolerance of ± 2 mil might be acceptable in the 4,000-mc cavity, this tolerance now becomes $\pm\frac{2}{15}$ mil or ± 0.13 mil for the same relative accuracy in resonant frequency. Furthermore, Q is inversely proportional to the square root of frequency. Hence, the millimeter cavity would have at best an internal Q lower by a factor of $\sqrt{15}$ or roughly 4, provided its surface smoothness was relatively as good as that of the 4,000-mc cavity. In practice, this means that our millimeter cavity should have mirror smooth copper surfaces. This in turn, renders drawn and plated parts or, perhaps, machined cavities, which are perfectly acceptable at 4,000 mc, unsuitable for the millimeter band.

Next, consider the method of tuning which, in most internal cavity type tubes involves changing the grid separation in order to vary the effective shunt capacitance of the resonator. In a 4,000-mc tube this can be realized by making the top wall of the resonator a flexible membrane. Such a diaphragm, having a diameter of about one inch and containing properly dimensioned concentric corrugations and radial slots will withstand many thousands of tuning cycles. For practical reasons, such performance cannot be expected of a tuning diaphragm having a diameter of some 60 mils. This mechanical difficulty, then, forces us to look at modifications of the passive circuit which will permit the use of a sufficiently large tuning diaphragm.

* Because of the small dimensions involved, it is convenient to express dimensions in mils, i.e., in thousands of an inch.

The ratio of megacycles of frequency shift per mil of change in gap spacing is about 500/30 for a typical 4,000-mc reflex klystron such as the Western Electric 431A. For a cavity scaled up in frequency by 15 this factor will become 225 times greater or 3750 mc/mil. One mil change in gap spacing would therefore detune this hypothetical cavity by close to 4,000 mc. This, incidentally, makes the millimeter wave reflex klystron so sensitive an indicator of dimensional changes, that special measures must be taken to provide a tuner of adequate dispersion be it of the mechanical or thermal type. Steps must also be taken to prevent the transmission to the tuning diaphragm of undesired motions which might easily result in the destruction of the grids. Such motions might be caused by differential thermal expansion during processing on the pump and would be difficult to predict.

The electron gun poses another problem which cannot be solved by simple scaling. Since the total effective beam current should be at least equal to and preferably greater than that required for lower frequency tubes of medium power output, scaling down of the cathode surface would result in prohibitive current densities so far, at least, as oxide coated cathodes are concerned. This, then, requires an electron gun of the highest possible convergence. At the same time it is necessary to: (a) arrange for a mechanical structure capable of the high degree of precision in spacings and concentricity required for such guns, yet provide adequate thermal isolation for the cathode, and (b) devise grids which will withstand bombardment by an electron beam having a current density many times greater than that encountered in lower frequency klystrons.

Or, let us examine the problem of heat dissipation. Essentially the entire beam power which may be of the order of 25 watts is dissipated on the central cavity post. If this cavity post and the grid contained therein are to withstand this concentrated electron bombardment, these elements must be surrounded by a very effective heat sink which, in turn, must carry the heat to the outside of the tube envelope where either natural or forced convection and conduction cooling may be applied.

The performance of a millimeter wave klystron cannot be predicted on paper with any certainty. Too many are the approximations contained in the small signal theory and too many are the assumptions and guesses involved in estimating such important tube parameters as effective beam current, beam coupling coefficient, resonator shunt conductance, etc. At best, theory predicts that the tube should oscillate if the cavity Q is not degraded to a value greatly below that caused by frequency scaling and if the reflector space geometry gives rise to effi-

cient bunching. It may come as somewhat of a disappointment to many readers that the design of a millimeter wave reflex klystron in many respects is not a quantitative science. The philosophy, then, with which we approached this task was an essentially simple one. It called for an attack on the individual problems outlined earlier with a view to fitting the better solutions into the most promising overall pattern.

THE ELECTRON GUN

The design objectives for the electron gun were:

(a) high perveance,
(b) oxide coated, unipotential cathode,
(c) long life, which with (b) necessitates a highly convergent beam, and
(d) no accelerating grid.

The considerations leading to these objectives are readily discernible except, perhaps, (d); apart from mechanical complexity, the use of accelerating grids has, in the past, frequently given rise to instabilities associated with ion oscillations.

Two basic gun configurations, known to us at the time, seemed capable of further refinement and thereby of meeting the objectives listed above. One was the well known Pierce gun and the other one a gun developed by O. Heil[3] at Ohio State University.

Small guns of the Pierce type had been investigated during the war by E. M. Boone in connection with the 1464XQ. The best gun developed in the course of that study and the one used in the 1464XQ performed about as follows:

accelerating voltage	400 volts
cathode current	20 ma
cathode current density	425 ma/cm^2
perveance	2.5 microamps/volt$^{3/2}$
transmission efficiency through a 0.025″ dia. ungridded aperture	80 per cent
current density multiplication	12

The performance of the Heil gun as described in the literature looked so promising that a thorough study of its suitability for use in the M1805 seemed imperative. The published value of current density multiplication of 230 was almost 20 times greater than that of the Pierce gun described above and the perveance of 4.5 $\mu A/V^{3/2}$ almost twice as great. Accordingly, a number of sealed-off and gettered gun testers were built containing scaled-down Heil guns. In spite of careful scaling, however, and for reasons not entirely clear to us, these testers fell short of

duplicating published performance. Their perveance came fairly close to the published value but their transmission ranged around the unusably low value of 50 per cent. These initial results strongly suggested the need for extensive changes from the Heil gun as published if it was to play its intended role in the millimeter wave reflex klystron.

The basic outlines of the Heil gun are shown in Fig. 1. Its cathode is a portion of a spheroid formed by the rotation of an ellipse about its minor axis. The ratio of the major to minor axis, b/a, equals 1.295 and the height of the spheroid measured along the minor axis is 0.423 a. If the missing area of the semi-spheroid is assumed to be cylindrical (which is a very close approximation) the expression for the cathode area simplifies to,

$$\text{Cathode area (in cm}^2) = 7.85\, b^2 \tag{1}$$

where

b = major axis in inches

This equation is plotted in Fig. 1 along with cathode current density for

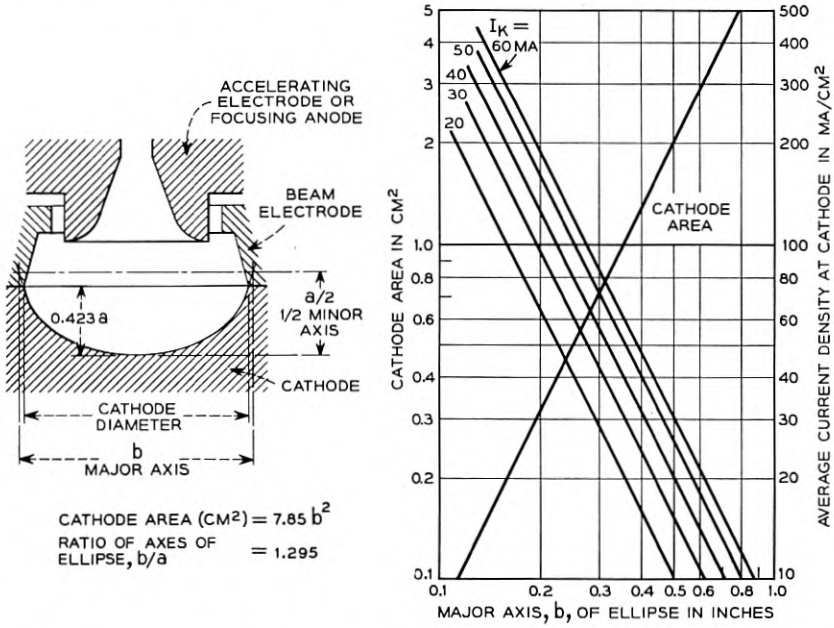


Fig. 1 — Design data for the Heil electron gun. The drawing to the left shows the basic configuration of the Heil gun and the plots on the right show the dependence of cathode area and average current density upon cathode size.

several values of current. The curves can be used to determine average current densities for proposed cathode diameters.

Closely adjacent to the cathode and joined to it electrically but isolated thermally is the beam forming electrode. Its shape as well as that of the accelerating electrode are also shown in Fig. 1. The angle subtended by the beam electrode and the tangent to the cathode at the point where it joins the beam electrode is 157.5°. This provides the well known 67.5° angle between the beam electrode and the trajectory of edge-electrons.

The program we undertook to explore the Heil gun for use in the millimeter wave klystron was a purely empirical one. It involved the systematic variation of the parameters d and f shown in Fig. 3. Dimension d, the spacing between cathode and focusing anode, controls the perveance of the gun and dimension f denotes the aperture diameter which in the final gun closely approximates the minimum beam diameter.

In order to provide for a vacuum environment closely resembling that typical of actual tubes, all gun testers were housed in sealed-off and gettered glass envelopes. Within these envelopes, however, the structures

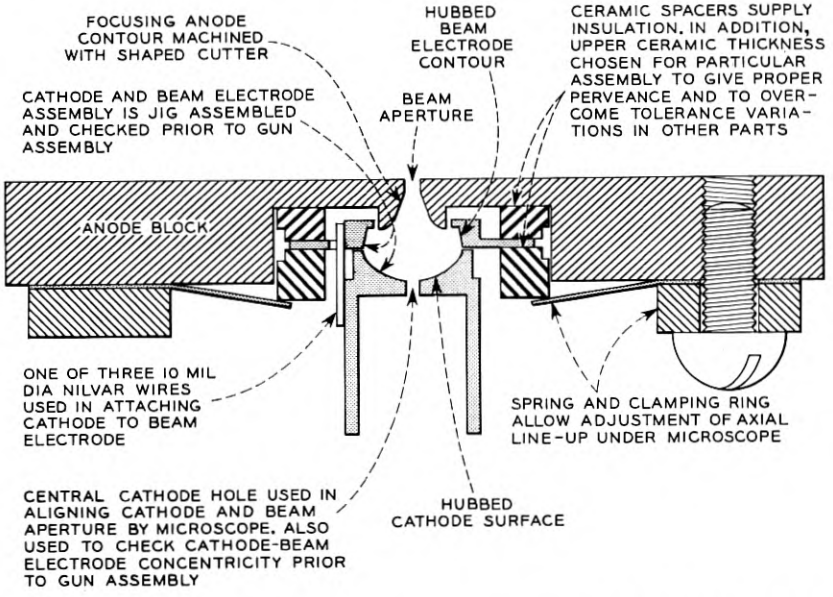


Fig. 2 — The demountable electron gun structure shown in this figure was originally used in an experimental study aimed at adapting the Heil gun for use in the M1805 millimeter wave reflex klystron. The reproducibility and mechanical stability, however, proved so satisfactory that this structure was incorporated in the final tube design essentially without change.

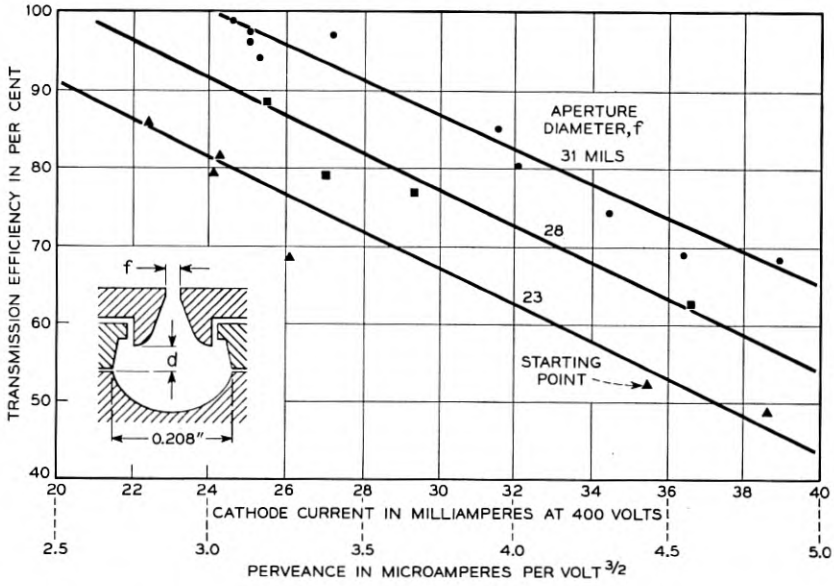


Fig. 3 — Experimental results obtained by varying the perveance and beam aperture of a Heil gun having a cathode scaled to 208-mil diameter. Each point represents the result obtained with a sealed-off and gettered gun tester. Perveance was changed by varying d. This, in turn, was achieved by the use of ceramic spacers (see Fig. 2) having different thicknesses thereby moving the cathode and beam-electrode as a unit. The diameter of the beam aperture, f, was enlarged in successive steps by cylindrical reaming.

were completely demountable as shown in Fig. 2. The gun testers consisted of three major elements: the collector assembly comprising the collector proper and secondary electron shield (not shown in Fig. 2), the cathode beam-electrode assembly and the anode block to which the first two assemblies were bolted. To achieve consistent and reproducible performance, it was found that dimensions had to be controlled to a degree which could not be achieved with self aligning parts. The following techniques were therefore adopted.

The emitting surface of the cathode was produced from a solid nickel blank with great accuracy by hubbing, a technique which will be described in greater detail in connection with the passive circuit. Other, less critical, cathode contours were machined. Three Nilvar wires, dimensioned for good thermal insulation consistent with adequate mechanical rigidity, were welded to both cathode and beam electrode, thereby yielding an assembly which could be carefully inspected both for concentricity and spacing. By means of two ceramic spacers and a

combination of spring and clamping ring, this assembly was electrically insulated from and properly spaced with respect to the anode block. The lower ceramic served as insulator only and its dimensions were therefore quite uncritical. The upper ceramic determined the cathode-anode spacing and hence had to have a closely controlled thickness and truly parallel surfaces. Instead, however, of keeping its thickness to a fixed value we found it preferable to maintain a stockpile of spacers having known and graded thicknesses and to select one which gave the desired gun perveance thereby overcoming dimensional variations in other parts.

The spring and clamping ring combination, in conjunction with the microscopic alignment procedure to be described, made concentricities of the order of ¼ mil readily achievable. This procedure involved assembling the constituent gun parts as in Fig. 2 but with the clamping ring tightened just enough to hold them together yet permitting lateral motion of the cathode assembly with very light pressure. The assembly was then mounted on a monocular microscope equipped with a reticule consisting of concentric circles. By alternately focusing on the beam aperture and the central cathode hole — which, incidentally, was provided for this purpose only and has no electrical significance — the cathode assembly could be brought into essentially perfect alignment and the clamping ring tightened.

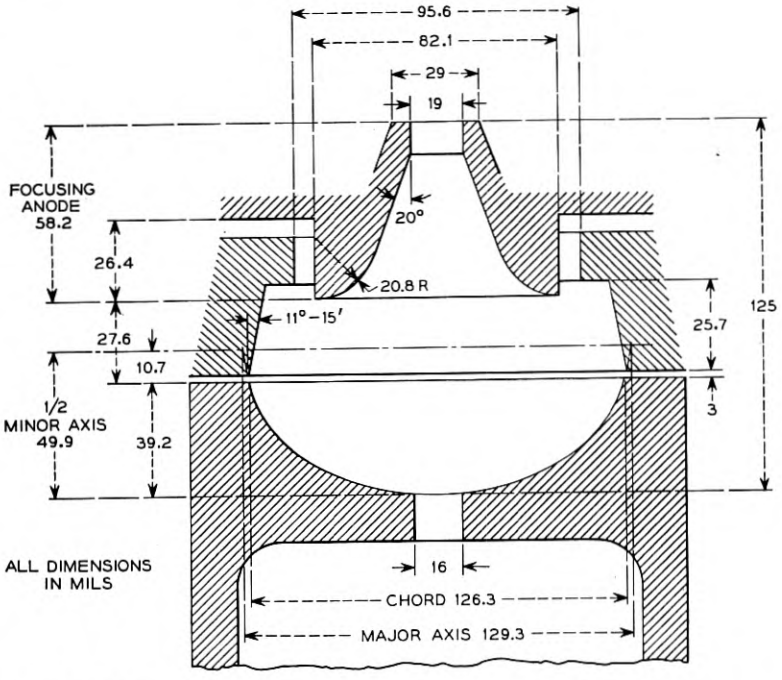
There are two simple yet effective means of controlling the transmission efficiency of an electron gun; variation of the spacing between cathode and focusing anode — and hence of perveance — and adjustment of the limiting aperture size. In the Heil gun of Fig. 3 this corresponds to varying dimensions d and f respectively. As mentioned earlier, the starting point for this study was a scaled down version of the original Heil gun, scaled down to a cathode diameter of 208 mils, with a corresponding beam aperture of 23 mils dia. and housed in the semi-demountable structure described earlier. The experimental result obtained with this tester after seal-off and under space charge limited conditions is represented by the point marked "starting point" in Fig. 3. We see that its perveance equalled about 4.4 $\mu a/V^{3/2}$ and was therefore approximately correct but that its transmission efficiency of about 50 per cent was too low to be useful. Keeping the diameter of the beam aperture fixed at 23 mils and varying the gun perveance by substituting upper ceramic spacers (see Fig. 2) of different thicknesses, F. P. Drechsler obtained the triangular points of Fig. 3 which define the lowest of the three straight lines shown. This particular series of experiments was stopped before the transmission efficiency had increased beyond 85

per cent since the perveance had then dropped below 2.8 $\mu a/V^{3/2}$. Drechsler then enlarged the diameter of the beam aperture (dimension, *f*, Fig. 3) by cylindrical reaming and traversed the same range of perveances, obtaining the experimental points which define the middle curve of Fig. 3. Finally, the experiment was repeated with a 31 mil diameter aperature. This yielded a group of gun testers with a transmission efficiency clustered around 95 per cent and a perveance slightly greater than 3 $\mu A/V^{3/2}$. As this performance was considered satisfactory, the configuration giving rise to it was selected as the prototype for the millimeter tube gun. It merely remained to scale this design down by a factor of about $\frac{2}{3}$, that is, to a beam aperture of 19 mils, thereby making it compatible with the dimensions established in the course of the cavity study to be described later. The significant dimensions of this modified Heil gun of final size and its performance under various operating conditions are given in Fig. 4. It should be noted that the tabulated values of cathode current and cathode loading were obtained with a unidirectional electron beam such as is normally produced in gun testers. In an operating reflex klystron a certain fraction of electrons are returned to the cathode region and perveance will therefore be somewhat lower.

The mechanical gun structure described earlier was conceived originally as a convenient yet precise vehicle for gun studies. So favorable, however, has been our experience with this method of gun mounting, in ease of assembly, in reproducibility of results and in thermal stability, that it was incorporated in the M1805 essentially without change.

The cathode currents of more than 90 per cent of the tubes built fell within 10 per cent of the mean, that is, within 2 ma of 18.7 ma when operated at 400 volts. About 95 per cent of these same tubes had transmission efficiencies within 5 per cent of the mean value of 56 per cent. The value of cathode current at a given beam voltage is a sensitive measure of the spacing between cathode and focusing anode (dimension, *d*, in Fig. 3). If we assume that this is the only source of the variation in perveance, the experimental data indicate that this spacing is being held to within 1.5 mil for the great majority of tubes. The value of transmission efficiency as such has little significance since secondary electrons are not controllable in the tube structure. However, the spread in values of transmission efficiency is closely related to the reproducibility in gun alignment. To obtain the value of transmission efficiency experimentally, we must operate the repeller positively and measure the fraction of cathode current reaching it after having passed through two grids. In a normal reflex klystron, operation with a positive repeller would soon result in overheating and the consequent destruction of the tube. To

overcome this difficulty a directly-reading transmission measuring set was developed by C. L. Nenninger in which the electron beam was pulsed at a sufficiently low duty cycle. The information was presented as an oscilloscope display consisting of two pulses with heights proportional to cathode current and transmitted current respectively as shown in Fig. 5.



TRANSMISSION EFFICIENCY ---------- 95 %

PERVEANCE ---------- 3 MICROAMP/VOLT$^{3/2}$

CURRENT DENSITY MULTIPLICATION ---------- 75

HEATER POWER ---------- 3.9 WATTS

BEAM VOLTAGE (VOLTS)	400	500	600
CATHODE CURRENT (MA)	24	33	45
CURRENT DENSITY AT CATHODE (MA/CM^2)	180	250	340
CURRENT DENSITY AT BEAM MINIMUM (AMPS/CM^2)	14	19	25

Fig. 4 — Modified Heil electron gun as used in M1805 millimeter wave reflex klystron. The performance data tabulated above, only apply to a unidirectional electron beam as produced in gun testers. Both perveance and cathode loading will be somewhat lower in an operating reflex klystron due to electrons returned into cathode region.

THE PASSIVE CIRCUIT

Of the two most commonly used mechanical means of tuning a klystron cavity, namely, capacitive and inductive tuning, tubes of the internal cavity type almost exclusively employ capacitive tuning. Generally, this method of tuning involves changing the separation between the interaction gap grids; this changes the effective shunt capacitance and hence the resonant frequency of the cavity. For a given frequency change it requires a much smaller physical motion than inductive tuning and one which is more readily realized mechanically. Capacitive tuning, therefore, became the obvious choice for the M1805.

How did this decision affect the design of the passive circuit? Lower frequency reflex klystrons are usually tuned by distorting a metal diaphragm which might be the cavity wall containing one of the interaction gap grids. Such an elementary approach, when applied to a millimeter wave cavity, would lead to many practical difficulties. Fig. 6(a) shows the shape and size of the frequency-determining, resonant cavity. Its upper surface is seen to have a diameter of only 76 mils, clearly too small to afford the flexibility required of a tuning diaphragm; this problem is further aggravated by the difficulty of scaling such items as tuning links or the size of brazing fillets. The solution, here, was to surround the inner, frequency-determining cavity by a stepped, radial choke section which presents a good broadband short to the inner cavity. As shown in Fig. 6(b) the diaphragm now extends across both the cavity and choke section, the latter having a diameter of 284 mils, almost four times that of

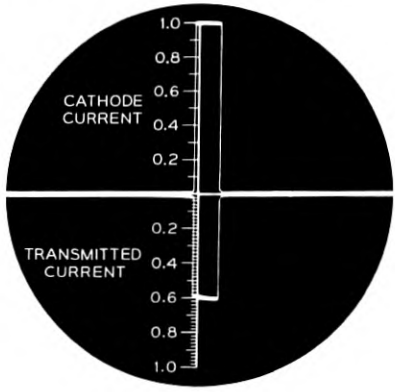


Fig. 5 — Oscilloscope display obtained with electron-gun-transmission test set. The special reticule permits the cathode current pulse to be adjusted for unit height; thus the height of the transmitted current pulse gives transmission efficiency directly.

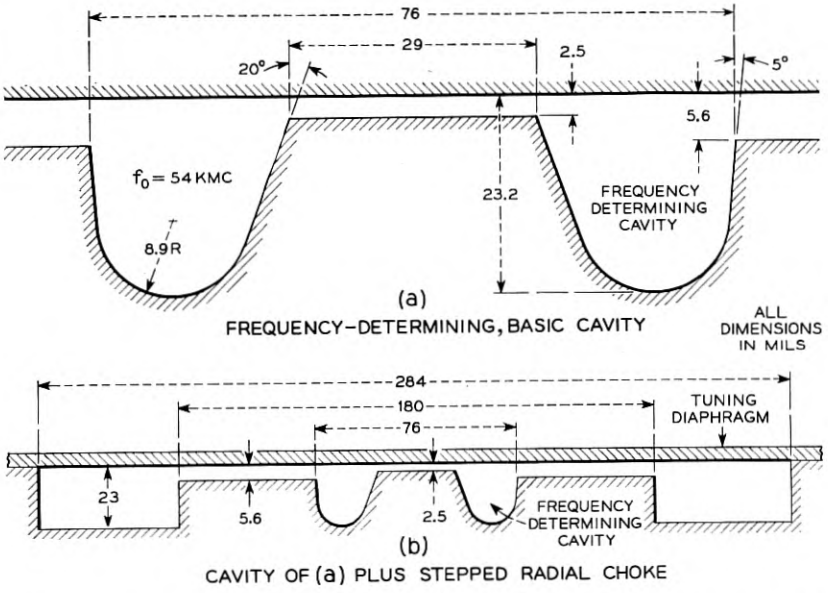


Fig. 6 — Resonant cavity of M1805 millimeter wave reflex klystron — (a) the shape and size of the inner, frequency-determining cavity and (b) the same cavity surrounded by a stepped, radial choke section. The choke section presents a broadband short to cavity proper, yet permits the use of a tuning diaphragm sufficiently large for flexibility. The resonant frequency of the cavity with the gap spacing as shown is about 54 kmc. This frequency is not appreciably affected by the addition of the choke section.

the primary cavity. The exact configuration was established experimentally with the aid of machined brass models scaled for operation in the 4,000-mc band, a frequency range in which broadband waveguide components and adequate measurement techniques were available. The addition of the radial choke section was found to have a negligible effect on either the internal Q or the resonant frequency of the inner cavity. It did, however, complicate the design of a broadband output transformer which forms the connecting link between the resonator and the output waveguide.

For the operating reflex klystron to deliver maximum power to a matched waveguide without an external susceptance transformer, the external load must be coupled to the resonator with sensibly uniform tightness over the projected mechanical tuning range. In other words, the output transformer must give rise to an external Q, Q_E, which does not change appreciably with frequency. In the case of the resonator and choke system of Fig. 6(b), this must be achieved in the presence of

variations in height of the inner choke section. These variations of height and consequently of the characteristic impedance of the radial transmission line adjacent to the inner cavity are caused by the tuning motion of the diaphragm.

Some of the output coupling configurations investigated in the process of evolving a design which would have the required electrical properties yet be amenable to fabrication are shown in Fig. 7. Shown beside each configuration is an experimental plot giving the variation in external Q, normalized with respect to the midband value, over the frequency range of interest. Like the cavity study proper, this work too was performed in the 4,000-mc band using large scale brass models. It is seen, for instance, that the electrically simplest case, namely that of an iris opening directly into the output waveguide, Fig. 7(a), gives rise to a variation in external Q greater than 10:1, an obviously undesirable condition. The steep slope of this curve is primarily due to variations in height of the inner choke section. Thus, at the high frequency end where the external Q has its lowest value and the tube is consequently most strongly coupled to the external load, the inner choke section has its maximum height. As we tune to lower frequencies by closing the interaction gap, the inner choke height decreases and along with it the tightness of coupling; Q_E, therefore, increases. Proceeding to Fig. 7(b) it is seen that the interposition of a linear taper between iris and waveguide has reduced the variation in external Q from a value exceeding 10:1 to about 6:1. The use of a quarter-wave transformer as in Fig. 7(c) reduces the variation to less than 2:1.

It is conceivable that the frequency sensitivity of the quarter-wave transformer of Fig. 7(c) could have been adjusted to yield an essentially uniform external Q. It was deemed desirable, however, to eliminate the thin iris required for this design because of the serious structural difficulties it would have presented when scaled down by a factor of about 15. This consideration led to the output transformer of Fig. 7(d) in which the iris has been replaced by a quarter-wave transformer of full guide width brought right up to the outer choke section. This configuration not only is the mechanically simplest one and therefore well suited for scaling, but gave rise to the best electrical performance. It became the obvious choice for the final design of the M1805. The electrical performance of this transformer is determined by the length, L, and the height, H, (see Fig. 7d). Basically, the height controls the tightness of coupling and the length the frequency sensitivity.

The effect of the length, L, of the quarter-wave transformer on the frequency sensitivity of external Q is shown in the experimental curves

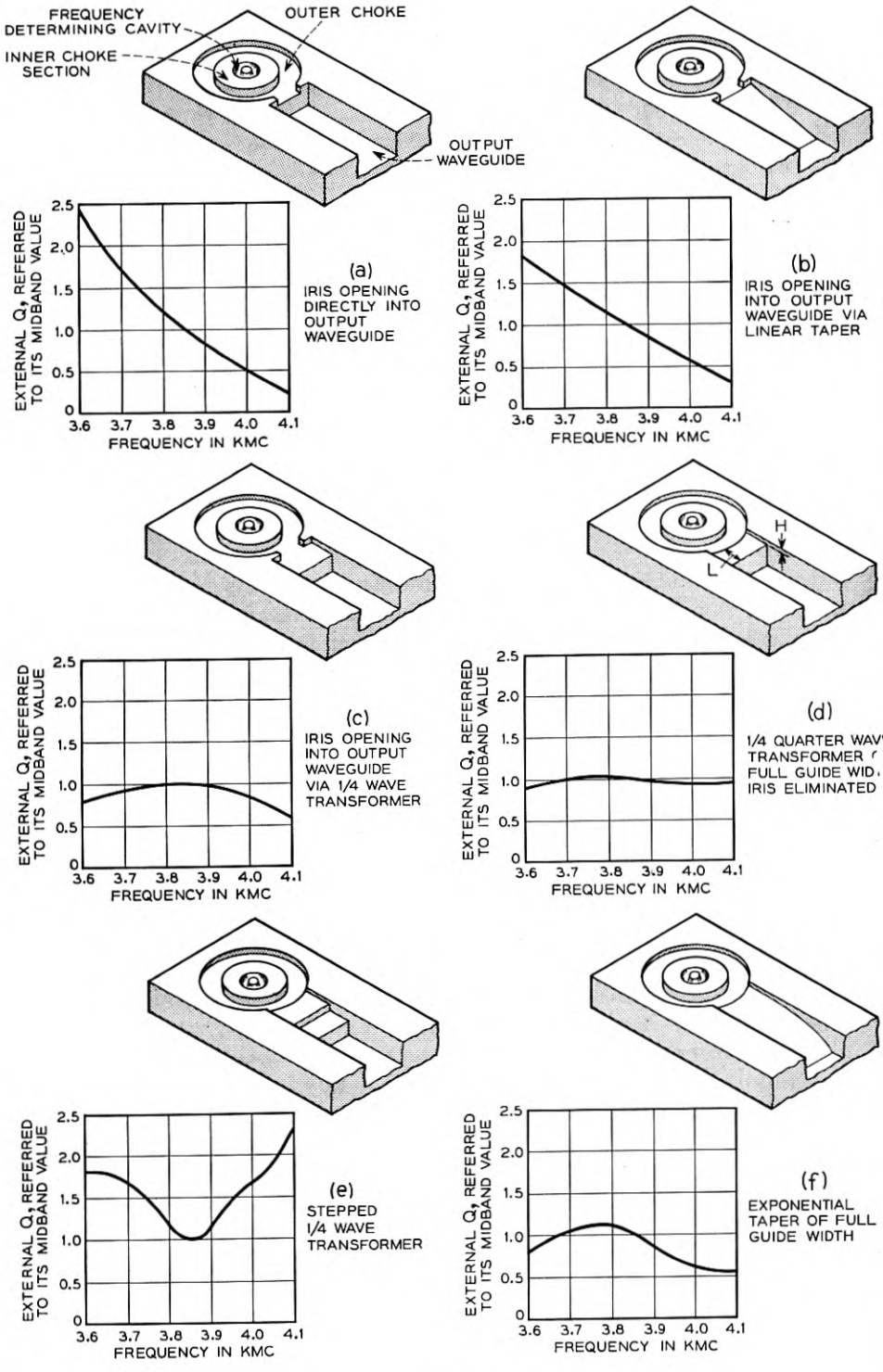

FREQUENCY DETERMINING CAVITY
INNER CHOKE SECTION
OUTER CHOKE
OUTPUT WAVEGUIDE
EXTERNAL Q, REFERRED TO ITS MIDBAND VALUE
2.5
2.0
1.5
1.0
0.5
0
3.6 3.7 3.8 3.9 4.0 4.1
FREQUENCY IN KMC
(a)
IRIS OPENING DIRECTLY INTO OUTPUT WAVEGUIDE
(b)
IRIS OPENING INTO OUTPUT WAVEGUIDE VIA LINEAR TAPER
H
L
(c)
IRIS OPENING INTO OUTPUT WAVEGUIDE VIA 1/4 WAVE TRANSFORMER
(d)
1/4 QUARTER WAV
TRANSFORMER
FULL GUIDE WID
IRIS ELIMINATED
(e)
STEPPED 1/4 WAVE TRANSFORMER
(f)
EXPONENTIAL TAPER OF FULL GUIDE WIDTH

FIG. 7

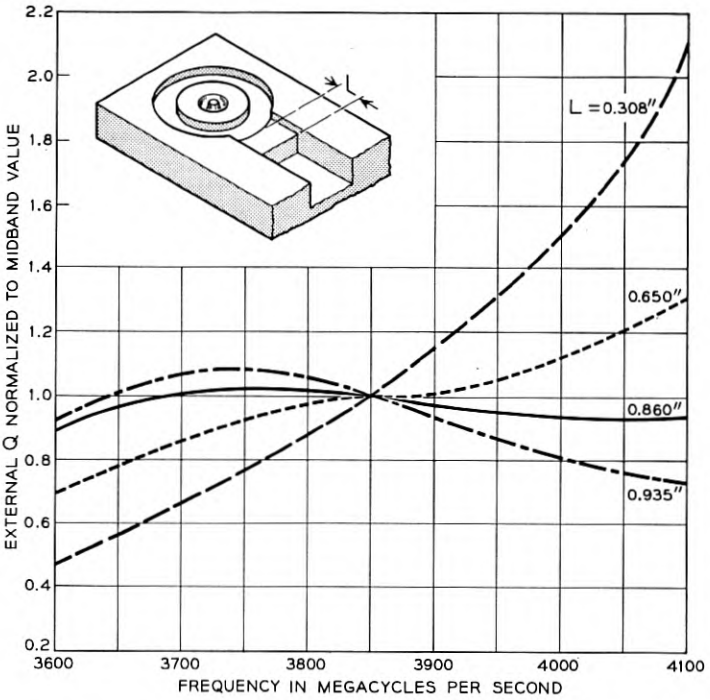


Fig. 8 — Variation in cavity loading (i.e., external Q) for various lengths, L, of the output transformer. The curve for $L = 0.860''$ shows the least variation and was chosen for M1805 design. The above curves were obtained experimentally using scaled-up, 4,000-mc brass models.

of Fig. 8. As shown by the solid curve in this figure it is possible, by a proper choice of L, to limit the variation in external Q to less than ± 3 per cent. We further see that not only can L be adjusted for maximum uniformity in loading but that its value might be chosen to produce controlled variations in external Q. These, in turn, could compensate for such variations in internal cavity losses as are shown in the curves of Fig. 9. The first two of these curves, namely, the ones relating internal Q and equivalent gap capacitance with frequency, were obtained by independent experiments using scaled-up brass models of the passive circuit and the third curve was calculated from the relation, $R = Q_0/\omega C$.

Fig. 7 — Various output coupling configurations and corresponding variations in external Q with frequency. The values of external Q have been normalized with respect to the midband value. The output circuit, (*d*), shows the least variation in external Q, only $\pm$ 3 per cent, and is the simplest mechanically. It was chosen for the final design of the M1805. Dimension "L" controls the frequency sensitivity of coupling and "H" its tightness. This work was performed in the 4,000-mc band using scaled-up brass models.

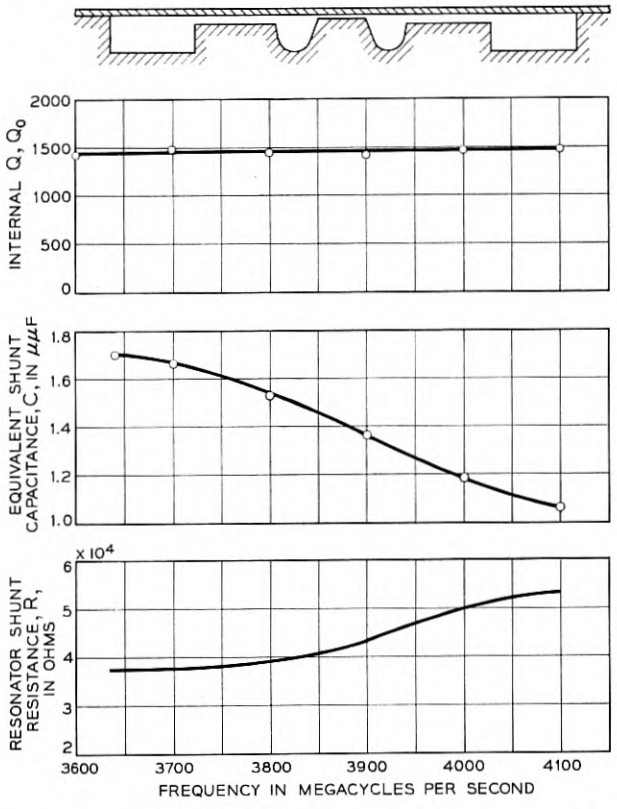


Fig. 9 — Variation of internal cavity parameters over the mechanical tuning range. Using a scaled-up brass model of the cavity shown in Fig. 6 (b), the curves for internal Q, Q_0, and for the equivalent gap capacitance, C, were determined experimentally by independent methods and the third curve calculated from the first two using the relation, $R = Q_0/\omega C$.

Disregarding the absolute values of shunt resistance because of the uncertainty of the applicable scaling factor but considering only its relative variation over the mechanical tuning range, we may well argue that a constant external Q will not extract the maximum available power at all frequencies.[4] Instead, the curve of external Q should be shaped such that the tube is more heavily loaded towards the high frequency end. This argument is further complicated by the fact that the beam coupling coefficient and consequently the electronic admittance also change as the grid separation is varied and that this change occurs in a direction tending to counteract the variation of resonator shunt resistance shown in Fig. 9. On the whole, then, the requirement of constant external Q seems

to be a reasonable initial assumption and one that can only be improved upon by experimental equalization with the aid of operating millimeter tubes of near final design. At the same time, being able to change the loading characteristics in a controlled manner appears to be a definite advantage, although one which has not as yet been exploited in the M1805. Performance data to be given later indicate that this tube delivers maximum, or near maximum, power into a matched waveguide with an output transformer scaled from the one corresponding to the solid curve in Fig. 8.

The objective of the model studies just described was a design which not only would have the desired electrical properties but would make possible a one-piece, hubbed cavity block. Accordingly, we have arranged that all the elements constituting the passive circuit extend *into* the cavity block from a common and plane surface and, in addition, we have eliminated the thin-walled iris which would normally connect the resonator with the output waveguide. This one-piece, hubbed structure has given rise to excellent reproducibility despite the minute sizes.

The method of fabrication of the cavity block as evolved by L. B. Luckner, proceeds as indicated in Fig. 10. The starting point is a simple cylindrical blank as in Fig. 10(a) made of vacuum melted, OFHC copper and having a polished top surface. A hardened and highly polished steel die, the inverse in shape of the impression to be hubbed, is forced into the cold blank by means of a hydraulic press with a thrust of about 16 tons. The resulting impression of the resonant cavity and output circuit has a very dense and essentially mirror smooth surface. To prevent any accidental marring or scratching of this surface and to preserve the finish during subsequent machining operations, the impression is filled with resin immediately after it has been hubbed as shown in Fig. 10(b). By means of three reference holes impressed into the blank during the original hubbing, the cavity block is accurately chucked in the lathe where the focusing anode and the adjacent contours are machined. All that remains now is the relatively straight-forward machining of the common flat surface to give the finished block as shown in Fig. 10(c).

The finished cavity block may be seen more clearly in the cutaway drawing of Fig. 11. Contained in this single unit are the following essential tube elements:

1. Inner, frequency determining cavity
2. Stepped radial choke section
3. Broadband output transformer
4. Output waveguide
5. Focusing anode for Heil gun

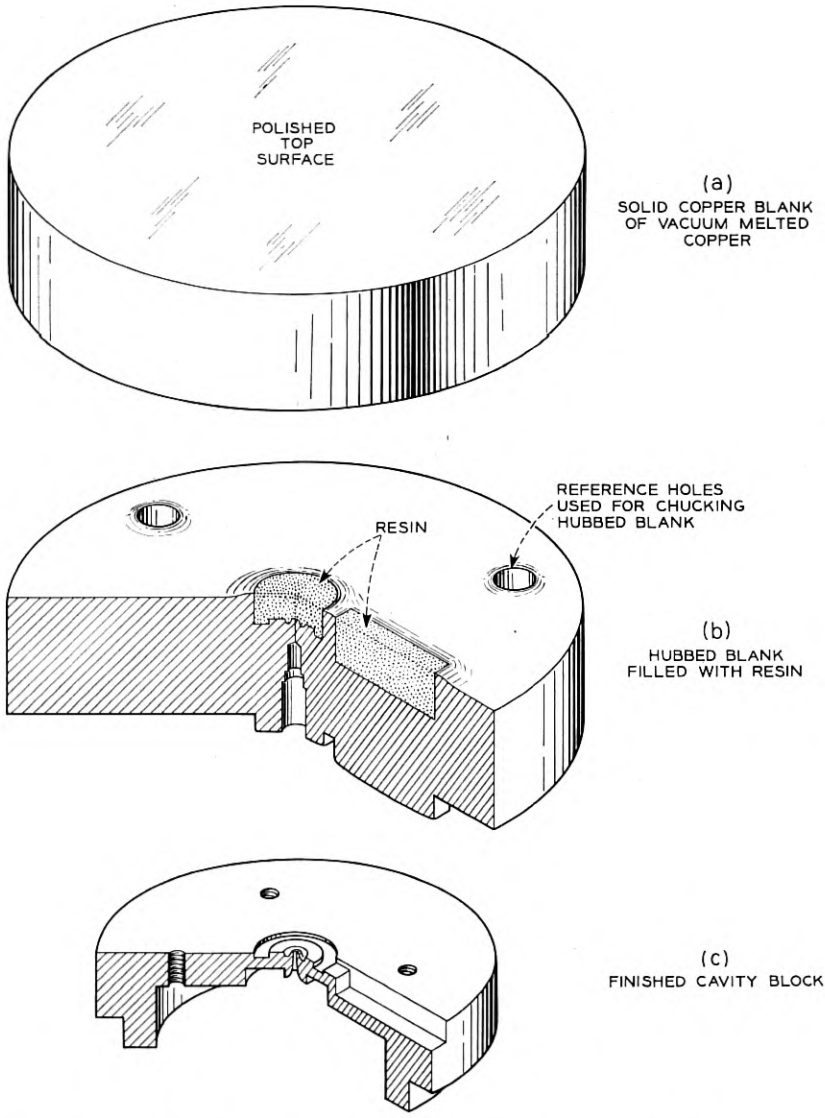


Fig. 10 — Principal steps in preparation of cavity block. (a) Starting point is a cylindrical blank made of vacuum melted, OFHC copper. (b) Copper blank into which has been hubbed the passive circuit and three reference holes used for precise chucking in subsequent machining. Resin filler prevents marring of mirror smooth surfaces. (c) Finished cavity block.

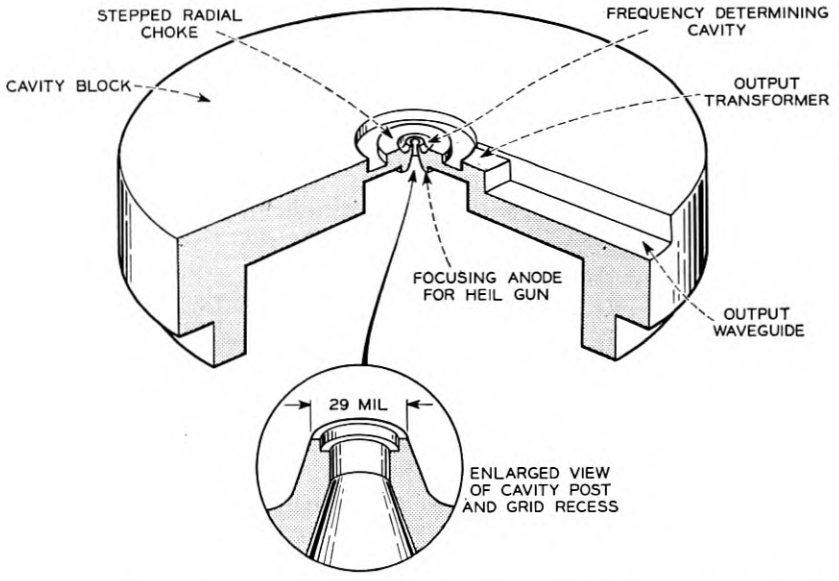


Fig. 11 — Cutaway view of completed cavity block. Made from a solid copper blank by the process outlined in Fig. 10, this unit really constitutes the heart of the tube.

In addition, we have hubbed a recess into the cavity post (see enlargement in Fig. 11) to serve as a nest for one of the interaction-gap grids.

By this design of the passive circuit, a number of significant electromechanical advantages have been achieved. These are:

1. All surfaces carrying high frequency currents are produced solely by hubbing. They require no further machining or handling so that their mirror like surface finish may be readily retained.

2. The hubbing process and the one-piece construction have eliminated the normal build up tolerances and thereby given rise to excellent reproducibility.

3. By rounding the contour of the inner frequency-determining cavity a favorable shape factor has been achieved.

4. The lossy braze joint between diaphragm and cavity is located in the outer choke section where it does not appreciably degrade the internal Q.

5. Since the copper block extends to the outside of the tube envelope, forced air cooling may be applied effectively and a comparatively cool heat sink presented to the central cavity post and the grid it contains.

Since at 60,000 mc the skin depth in copper equals about $\frac{1}{100}$ mil, surface layers of low conductivity have a very pronounced effect on cavity Q. Such lowered conductivity may be due not only to chemical impurities or surface roughness but may be the result of worked surface layers.

In a recent British study,[5, 6] the RF conductivities of hubbed copper surfaces at 8 mm wavelength were reported to be from 30 to 50 per cent lower than the dc conductivity. The same study also describes a number of surface treatments which, when applied to 8-mm hubbed cavities, yielded values of internal Q very close to the theoretical value. In general, these processes involved the relief of residual strain by annealing and the removal or coverage of surface layers by chemical means. Only the first of these techniques, namely annealing, has been applied to the M1805 cavity block to date. The application of the chemical surface treatments suggested in the British report is planned but will have to wait upon a refinement of Q-measuring techniques in the 5–6 mm range. Rough Q-measurements made on M1805 cavities indicate that the hubbing process followed by annealing gives rise to a value of Q close to that extrapolated from measurements on brass models operating in the 4,000-mc band.

GRIDS

The need for a gridded interaction gap arose as a direct consequence of our decision to aim at low voltage operation. Apart from considerations of general convenience, it was felt that a low voltage tube would afford economies in the cost of power supplies and in the maintenance of future systems such as to more than outweigh the difficulty of providing grids. This difficulty was due not only to the exceedingly small size of the grids, 20- and 30-mil internal diameter for G1* and G2* respectively, but also to the very intensive electron bombardment to which they would be subjected. Thus, for an accelerating voltage of 600 and a unidirectional beam, the current density at the plane of G1 would be about 25 amps/cm^2 and the corresponding power density 15 kilowatts/cm^2.

Early millimeter wave tubes were equipped with grids of the type shown in the micro-photographs of Fig. 12. Comparatively simple in construction, these grids consisted of a number of parallel tungsten wires, 0.8 mil in diameter, embedded and gold brazed into the surrounding copper. When used as G1, this type of grid was found to withstand continuous bombardment by a 600-volt beam. When used as G2, however,

* G1 denotes the interaction-gap-grid closer to the cathode and G2 the grid closer to repeller.

the central portion was invariably eroded away. The heating conditions for G2 are, of course, much more severe than those for G1. The latter is smaller in diameter, connected to a more effective heat sink, namely the copper cavity, and, presumably, is bombarded by the outgoing stream only. G2, on the other hand, has a diameter of 30 mils (as against 20 mils for G1), it is bombarded by both the outgoing and returning stream and is connected to a relatively poorer heat sink, a rather thin copper diaphragm. This type of grid, then, limited the maximum beam voltage for continuous operation to about 400 volts. It was therefore necessary to operate early tubes having round wire grids under pulsed conditions if the beam voltage was increased beyond that value.

Since grids of the size used in the M1805 are cooled mainly by conduction, their dissipative capacity can best be raised by increasing the cross sectional area of the laterals, provided this does not result in increased electron interception. A grid consisting of fine ribbons or vanes rather than wires of circular cross section would, therefore, be of great advantage. The electron interception of such a grid would be determined primarily by the thickness of the ribbons perpendicular to the electron flow whereas the cross sectional area available for heat conduction would be proportional also to its depth parallel to the electron flow. A grid of this type was indeed suggested by R. L. Vance[7] of Bell Telephone Labo-

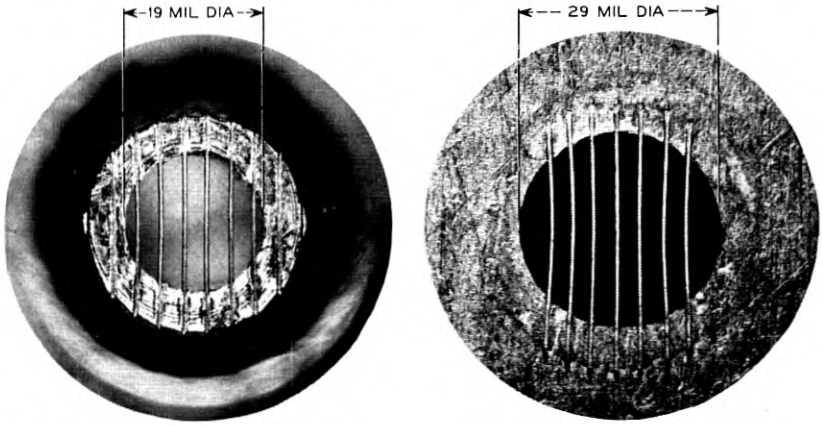


Fig. 12 — Round-wire tungsten grids used in early M1805 millimeter-wave klystrons. Consisting of several 0.8 mil diameter, gold plated tungsten wires embedded and brazed into the cavity post or diaphragm, these grids limited the beam voltage for continuous operation to about 400 volts. Higher beam voltages required pulsed operation.

ratories several years ago. Its basic configuration is shown in Fig. 13. Consisting of a number of tungsten vanes brazed individually into an outer cylindrical rim, this grid not only offers the increased heat dissipation capacity inherent in the use of vanes instead of wires, but the additional important advantage of well controlled and predictable thermal motions. These are results both of the cross sectional shape and the initial curvature which, in the presence of electron bombardment, will cause the "growing" grid laterals to continue in the direction of the initial curvature, i.e., without appreciably changing the gap spacing.

The special problem which confronted us was that of scaling this ribbon grid down to a size where Vance's suggested method of fabrication could no longer be used. We were greatly aided in the solution of this problem by engineers of the Sperry Gyroscope Corporation who very kindly disclosed to us the basic steps of an ingenious process by which ribbon grids with the basic configuration of Fig. 13 could be fabricated in very minute sizes. Important contributions to the detailed processing of these grids were subsequently made by W. Gronros, D. E. Koontz and F. P. Drechsler of the Laboratories.

The principal steps in the preparation of the ribbon grid are illustrated in Fig. 14. Copper plated ribbons of tungsten and iron — the tungsten 0.3-mil thick and the iron 3-mil thick — are wound and brazed into a tight spiral as shown in Fig. 14(a). Upon removal of the mandrel, the

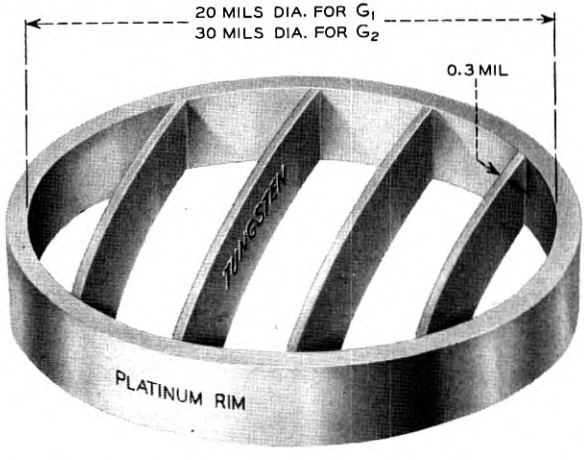


FIG. 13 — Sketch showing basic configuration of grid used in final version of M1805. The tungsten vanes, which present a thickness of 0.3 mil to the electron stream but are 3 to 4 mil deep parallel to the electron flow, are butt-brazed into the outer platinum rim. Their initial curvature serves to define the direction of expansion during operation at elevated temperatures.

brazed spiral is face ground on both sides to a thickness of perhaps 10 mil and chromium plated, the chromium serving as a thin barrier layer against subsequent brazes. The ground and plated spiral is then parted into several segments as shown in Fig. 14(b). Each segment is edge-ground to form a small, solid "pill" with a diameter corresponding to the inside diameter of the desired grid as indicated in Fig. 14(c). The grid pill is inserted and brazed into a short length of tightly fitting, thin walled platinum tubing which, for ease of handling, has previously been brazed into an iron disc; this is shown in Fig. 14(d). This entire assembly is then machine-lapped on both sides as in Fig. 13(e) to a thickness of 3 to 4 mils, i.e., to a thickness equal to the desired depth of the final grid. It merely remains to remove the iron both from outside and inside the platinum by etching. In the early stages of development, the etchant used was hot concentrated hydrochloric acid. Grids made in this manner, however, were not satisfactory because the acid did not completely remove the metallic deposits from the laterals. In addition to excessive electron interception, there was danger of these etching residues vaporizing at the high operating temperature. Deposition on the smooth cavity walls would cause increased RF losses while deposition on the insulation would cause electrical leakage. Moreover, hydrochloric acid also tended to attack the brazing fillets and thereby often gave rise to loose laterals.

A much more elegant process of iron removal was evolved by D. E. Koontz of the Chemical Department of the Laboratories. This process has consistently resulted in a high yield of ribbon grids of excellent quality. Briefly, it operates as follows: if iron is placed in a solution of aqueous copper chloride, the surface layer of iron goes into solution and is replaced by metallic copper. Ordinarily this reaction stops when the surface has been completely covered. The only way of maintaining this reaction is to continuously remove the deposited copper in order to expose the underlying metal to further attack. In Koontz's process, this is achieved by immersing the specimen of Fig. 14(c) in an ultrasonically agitated copper-chloride solution to which has been added a quantity of carborundum powder. The action of the suspended carborundum powder is twofold. It grinds off the surface layer of copper, as it is being formed, thereby providing the necessary condition for the complete dissolution of iron and it is quite effective in freeing the ribbon grid from any burrs which may have been raised in preceding lapping and grinding operations.

Since grids of the size used in the M1805 are cooled primarily by conduction, we may wonder whether the substitution of copper for tungsten would lead to a ribbon grid having adqeuate power handling capacity

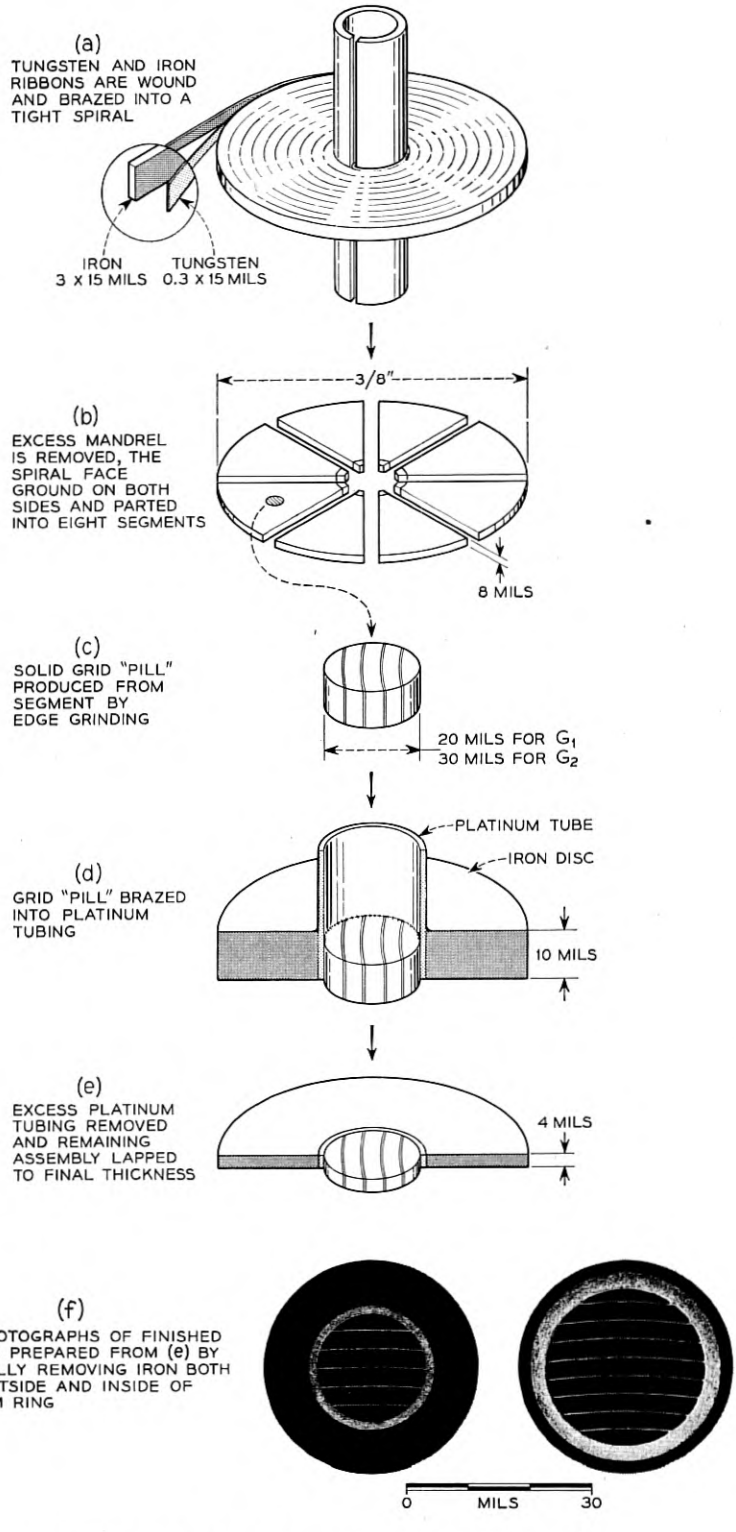


Fig. 14 — Basic steps in fabrication of ribbon grid.

yet with greatly reduced RF losses. The superior heat conductivity of copper might be thought to more than offset the lower melting point. This problem as well as others aimed at better understanding the nature of grid heating in millimeter wave klystrons were investigated in the course of a theoretical study by R. L. Wigington. He was able to show that a copper ribbon grid should operate much cooler than the equivalent tungsten ribbon grid but — for the case of G2 and a 600-volt beam — at a temperature too close to its own boiling point* to be useful.

Since the boiling point of tungsten is less than its melting temperature at normal tube pressures, overheating will first result in the erosion rather than the melting of the grid. Also, for the reasons already stated, the greatest possibility of burnout occurs at the center of G2. Fig. 15 shows theoretical plots of the maximum grid temperature, i.e., of the temperature at the center of the central G2-lateral, as a function of beam power for three types of grids. The first one of these grids, namely, the round wire tungsten grid, seems to leave the region of safe operation at about 400 volts and this agrees fairly well with experimental evidence. The second type of grid shown in Fig. 15, the tungsten ribbon grid, operates much cooler, staying well within the safe region at a beam voltage of 600 volts. Here, however, experiment indicates theory to be somewhat optimistic. This disagreement is, perhaps, not too surprising since we assumed parallel electron flow in calculating temperature distribution; thus the fractional electron interception depended on the projected grid area whereas the effective area is probably considerably greater. The expected performance of a copper ribbon grid is shown in the third plot of Fig. 15. Its calculated maximum temperature with a 600-volt beam is seen to come dangerously close to its boiling point. Since, for the reasons given earlier, the calculated temperatures are believed to be somewhat optimistic, this grid has not been constructed.

THE REFLECTOR

The last remaining tube element of electrical significance, namely, the reflector is shown in Fig. 16. Scaled down from an empirically derived design which had been found satisfactory in a 4,000-mc reflex klystron with a comparable electron-optical system, this repeller is the first and only one used to date. We are by no means certain that it constitutes an optimum design although there are indications that this first choice was a rather fortunate one. Experimental work aimed at optimizing a re-

* The "boiling point" is taken as the temperature at which the vapor pressure of the metal equals that of its environment. At normal tube pressure the boiling of copper will be near 1000°K and that of tungsten about 2500°K.

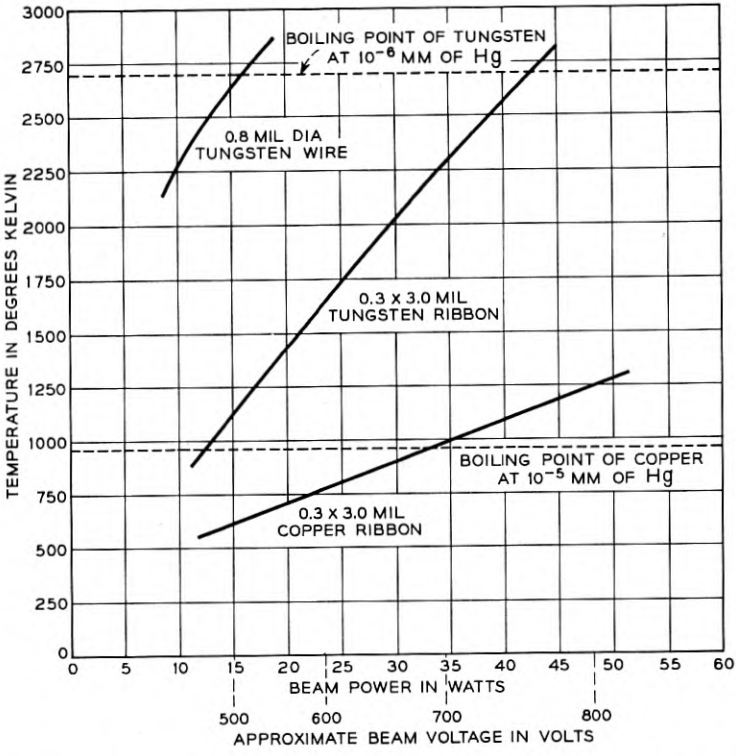


Fig. 15 — Calculated dependence of maximum grid temperature on the bombarding beam power for three types of grids. These curves have been computed for the specific geometry of the M1805. The maximum temperature occurs at the center of the central G2-lateral.

peller shape in our own laboratory and by other workers in this field have so far failed to disclose any sharp optima. There remains the important task, however, of demonstrating that the M1805 reflector design at least finds itself on a broad maximum.

Like the cathode, the electrically significant repeller contour is produced by hubbing and it too is aligned optically, with the help of a central aperture provided for this purpose only.

THE ELECTRICALLY SIGNIFICANT TUBE CONTOURS

Having covered all the essential tube components, individually, we may now proceed to show their interrelationship and their combination into a single overall pattern. Fig. 17 shows a scale drawing of the elec-

trically significant contours of the M1805 millimeter wave reflex klystron. It is interesting to note the large size of the cathode relative to the resonator and reflector regions. Note also that the centerline of the tube passes through four successive apertures, two of which, the cathode and reflector apertures, have been provided only to facilitate optical alignment.

One electrically significant component which has not been described in detail is the output window; it is a conventional choke-type glass window scaled into the millimeter band. The insertion loss of this window, the structure of which is shown in Fig. 19, ranges around 0.7 db and is fairly uniform over the band of interest.

CONSIDERATIONS IN CHOICE OF TUNER

It was stated earlier that this tube is tuned mechanically by changing the separation between the interaction gap grids. The total amount of motion required for tuning the tube from 50 to 60 kmc is less than four mils which corresponds to an average tuning rate of 0.4 micro-inch per megacycle. Fig. 18 shows the spread in tuning curves for seven tubes. Shown also (as the dashed curve) is the tuning curve which was predicted from the results obtained with large scale cavity models. In the

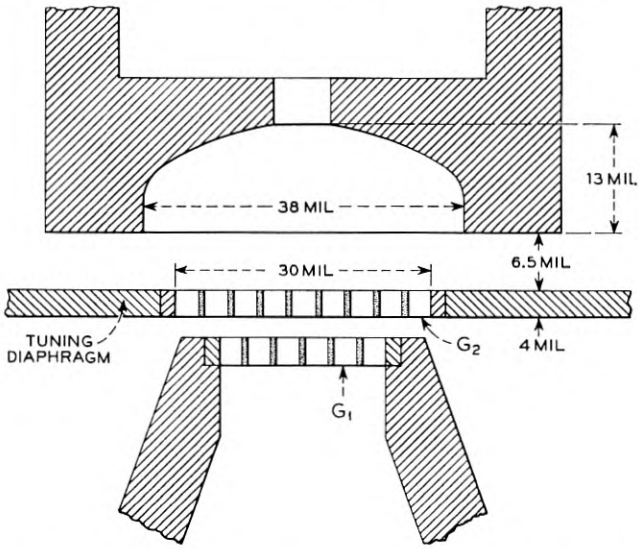


Fig. 16 — M1805, reflector space geometry. The central aperture in the reflector facilitates optical alignment.

face of the scaling factor of 15, such close agreement between model and actual tubes was rather gratifying.

Since the total tuning motion required in the M1805 is so small, we may wonder whether it would best be transmitted by mechanical or thermal means. Mechanical tuning would involve some type of linkage capable of producing very small motions in a controlled manner. In thermal tuning, we utilize motions induced by the thermal expansion of some member of a mechanical linkage, with the heating effect derived either from electron bombardment or from the passage of comparatively heavy electric currents. It is probably true that thermal tuning is inherently better suited to producing the delicate motions required in a millimeter wave tube. It must be remembered, however, that thermal tuning greatly complicates the structure inside the vacuum envelope and — external to the tube — invariably requires a feedback system to maintain frequency stability. Fortunately, though, the initial choice of tuner does not greatly affect the basic electrical design of the tube. For these reasons, we decided in favor of a mechanical tuner, a somewhat refined version of the tuner used in the 4,000-mc Western Electric 431A reflex klystron. The experience with this tuner has, on the whole, been

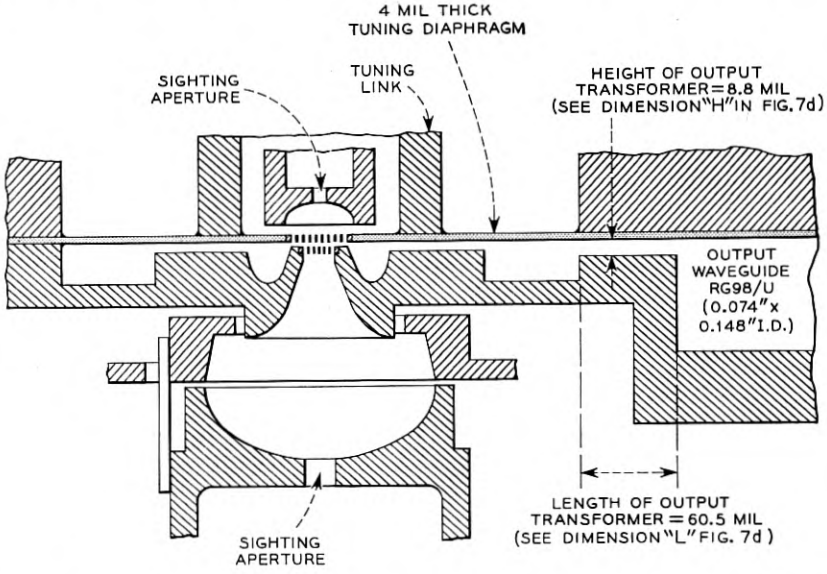


Fig. 17 — M1805, electrically significant contours. This drawing shows the interrelationship and the combination of the essential tube elements. More detailed data on these elements were contained in Fig. 4 for the gun, Fig. 6 for the passive circuit and Fig. 16 for the repeller.

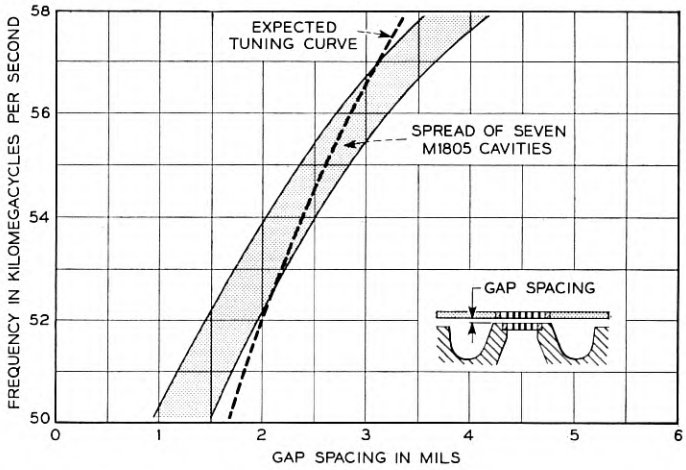


Fig. 18 — Cold tuning data for seven M1805 cavities. The dashed curve represents the tuning characteristic predicted for the millimeter wave tube on the basis of a 4000-mc cavity model. The difference in slope is probably the result of the lack of grids and, hence, the relatively greater gap capacitance of the low frequency model.

satisfactory. Just like the gun mounting structure described earlier it started out as a temporary expedient but, because of its satisfactory performance, was selected as our final solution. It is entirely possible, however, that future system applications may generate the need for a thermal tuner. Such a substitution could then be effected without the exploration of new techniques.

THE MECHANICAL STRUCTURE

The mechanical structure of the M1805 millimeter wave reflex klystron for which L. B. Luckner is largely responsible, is shown in the drawing of the completed tube of Fig. 19. In essence, it consists of three separate units which are combined by high-frequency brazes to form the finished tube. Proceeding from top to bottom, the first is the top bulb and tuner assembly. The second contains the entire passive circuit including the output window, the electron-optical system and — mounted on the cavity block by three struts — the stem and heater. The third element is a simple sleeve which slides over the stem and is brazed to it as well as to the cavity block to complete the vacuum envelope.

Referring to Fig. 19, we see that the mechanical tuner consists of a horizontal tuning arm pivoted approximately at its center on a vacuum

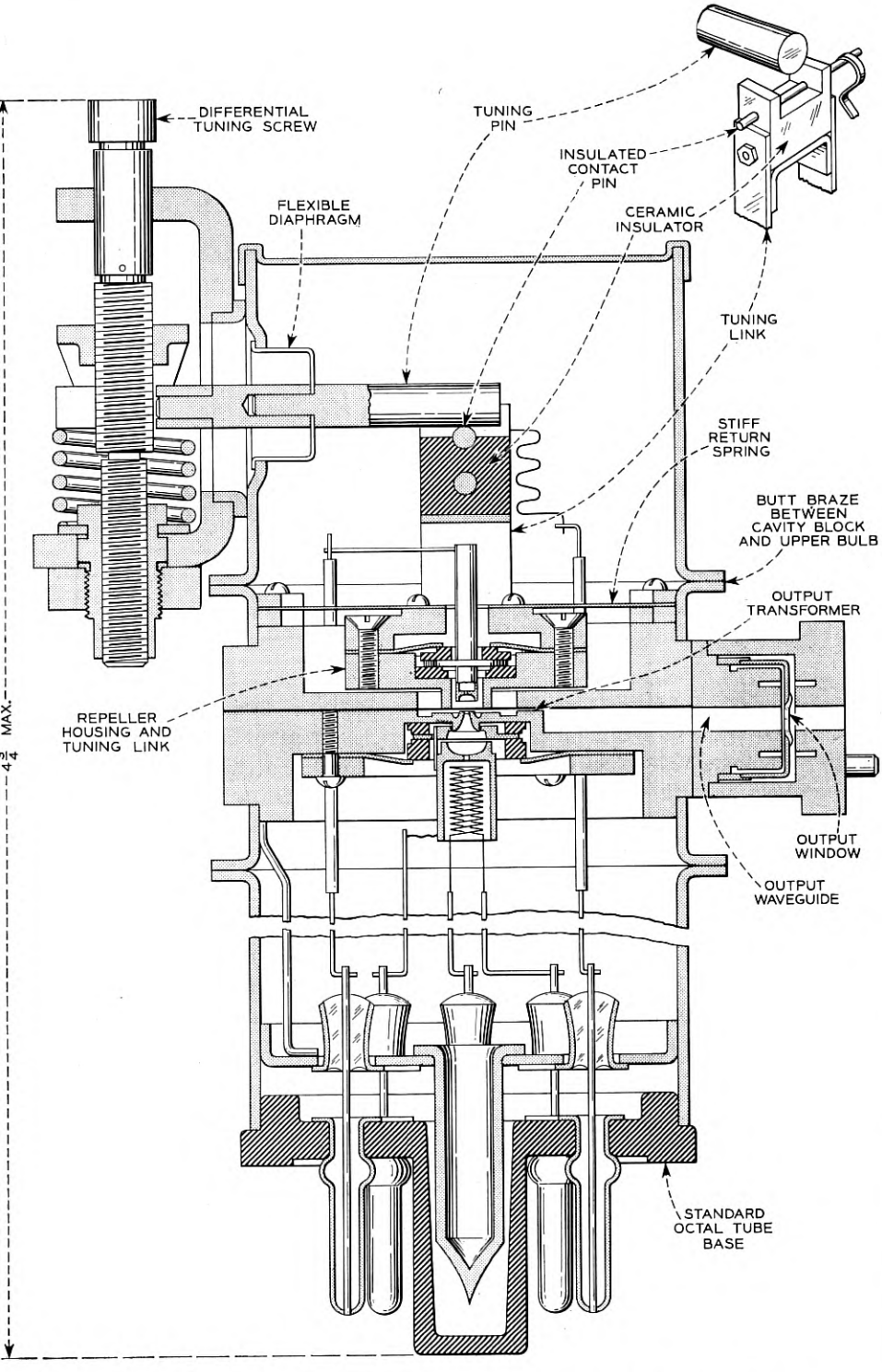


Fig. 19 — Layout of M1805, millimeter reflex klystron. The enlarged insert shows details of insulated contact pin which forms part of the alarm circuit.

tight diaphragm. Outside the vacuum, the tuning arm is acutated by a differential screw chosen such that one revolution moves the tuning arm some 0.3 mil along the center line of the tube.

The tuning arm is not rigidly connected to the tuning link but simply bears down on an insulated steel pin forming part of the tuning link. This tuning link is attached to the repeller housing which, in turn, is brazed to the tuning diaphragm. Hence, any motion of the insulated pin causes the equivalent motion of the G2-grid. The repeller housing and tuning link are firmly positioned by an Inconel-X return spring, the stiffness of which greatly exceeds that of the 4-mil copper tuning diaphragm. During pumping, this stiff return spring maintains an interaction gap spacing which corresponds to a resonant frequency somewhat higher than the upper limit of the mechanical tuning range. In the process of tuning, the tuning arm must simply exert sufficient downward pressure to overcome the restoring force of the return spring.

During bake-out and pumping, a spacing of several mils is maintained between the tuning arm and the insulated pin so as not to transmit undesired motion to the diaphragm. When the tube is ready for test, the tuning arm is brought down slowly until it contacts the insulated pin. The instant of contact is indicated by an external alarm circuit which may be connected between tube envelope and insulated pin; for this purpose a connection to the insulated pin is brought out to the base. From here on, any further motion of the tuning arm produces the corresponding motion of the G2-grid.

Another advantage of the external contact indication is that it makes possible an absolute frequency calibration such as is shown in Fig. 20(d). The tuning characteristic has been found sufficiently stable with time so that millimeter tubes can be rough-tuned to the desired frequency without the need for tedious wavemeter-tracking.

Referring to Fig. 19, we see that the cavity block can be reclaimed simply by unbrazing the upper and lower bulb. Since most of the cost of the tube is contained in the cavity block, this is an important feature and one which has demonstrated its usefulness on several occasions.

M1805, ELECTRICAL PERFORMANCE

Typical performance data are given in the experimental plots of Fig. 20. From Fig. 20(a) it is seen that the tube tunes mechanically from about 48,000 mc to 60,000 mc, that is, over a 12,000-mc band. This tube delivers a maximum of about 20 milliwatts of millimeter-wave power when operated with a 600-volt beam. Being approximately optimally coupled to a matched waveguide, a standing wave introducer is not re-

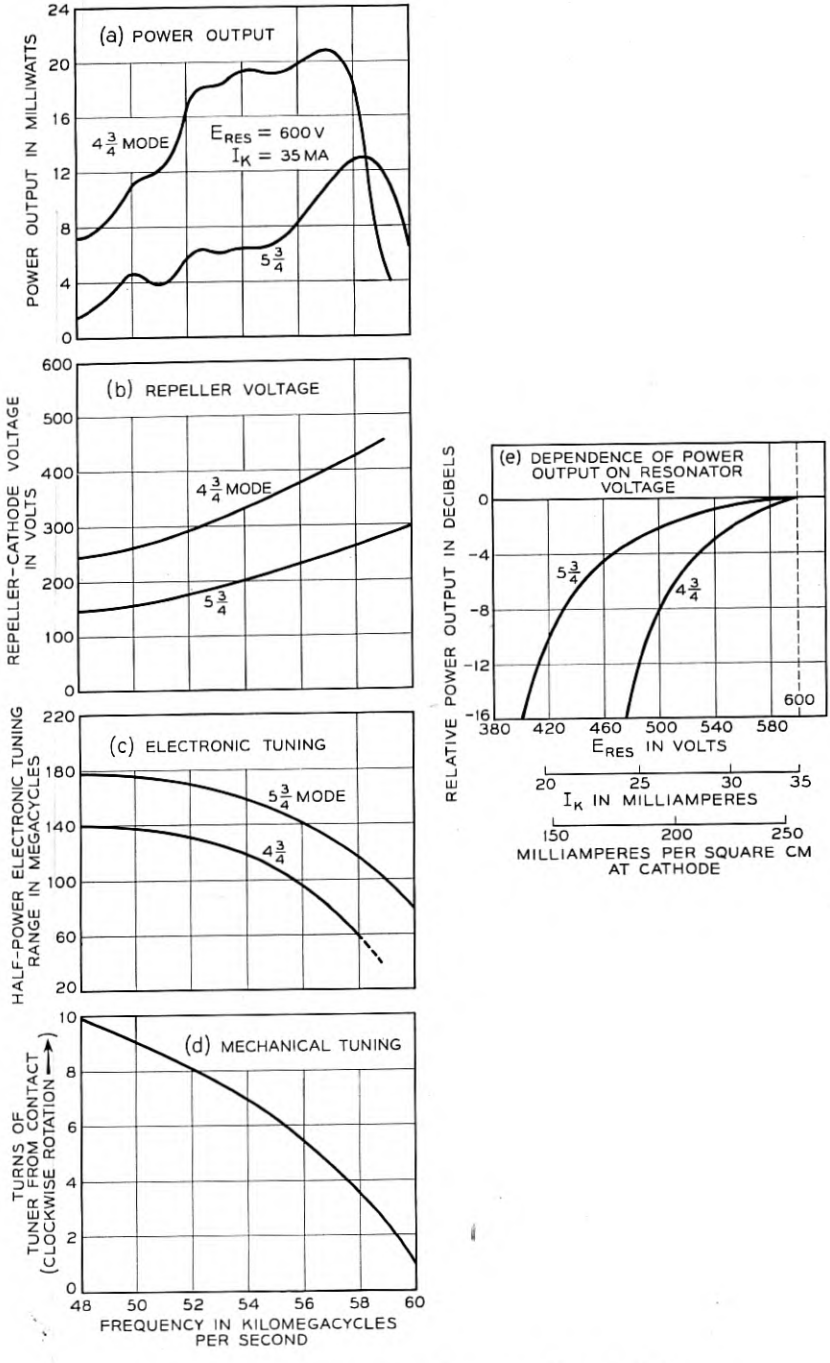


Fig. 20 — Typical M1805 performance characteristics.

quired for the extraction of maximum power. Individual tubes have delivered as much as 30 milliwatts of RF power, but the majority have ranged between 15 and 25 milliwatts. This spread in performance which includes data from early tubes has in part been traced to slight warpages in the diaphragm. Figures 20(b) and (c) relate the values of repeller voltage for maximum power output and the electronic tuning range with frequency for the $4\frac{3}{4}$ and $5\frac{3}{4}$ repeller modes.

The M1805 may be operated at beam voltages less than 600 volts, a typical dependence of power output on beam voltage being plotted in Fig. 20(e). If the beam voltage is decreased by 100 volts, that is, from 600 to 500 volts, the power output in the $4\frac{3}{4}$ mode drops about 8 db whereas it decreases by only 2 db in the $5\frac{3}{4}$ repeller mode.

The performance of the mechanical tuner is indicated by Fig. 20(d)

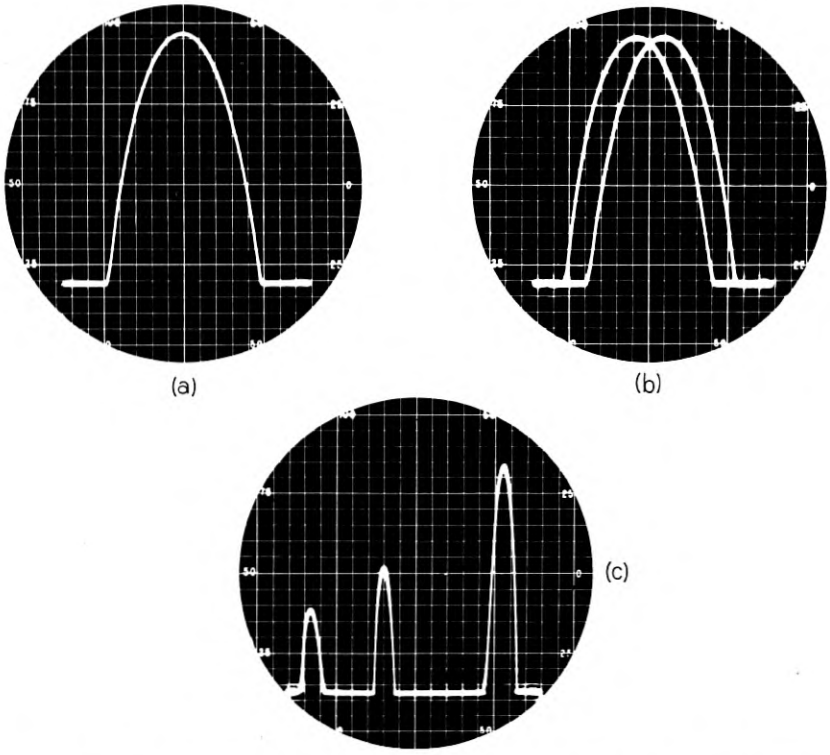


Fig. 21 — Oscillographs of typical repeller mode shapes obtained with sinusoidal sweep. In (a) and (c) the forward and reverse sweeps are coincident while they have been separated in (b) so as to facilitate examination of the individual traces. Oscillograph (c) shows the $4\frac{3}{4}$, $5\frac{3}{4}$ and $6\frac{3}{4}$ repeller modes.

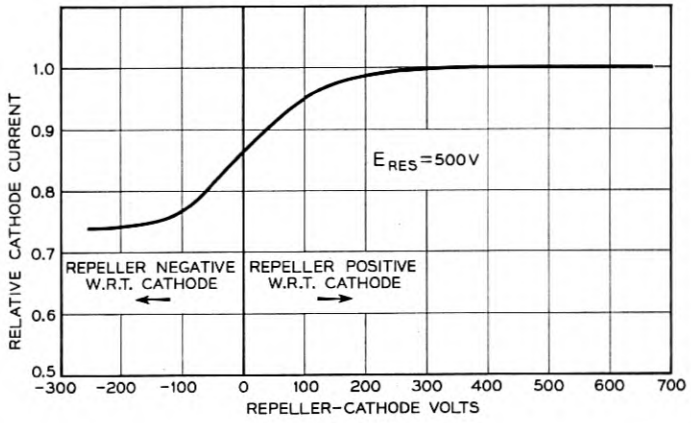


Fig. 22 — Variation of cathode current with repeller voltage. The values of cathode current have been normalized with respect to that obtained with a positive repeller. For the reflector space geometry of the M1805, operation with a negative repeller returns some electrons into the cathode region thereby causing a reduction in cathode current.

where the relation has been plotted between frequency and the rotational position of the tuning screw.

One noteworthy and quite consistent property in the performance of this tube has been the virtual absence of electronic hysteresis and the consequently very clean and symmetrical mode shapes. The oscillographs of Fig. 21 are typical. Oscillograph (a) shows a mode shape obtained with sinuosoidal repeller sweep and with both forward and reverse sweeps coincident. In oscillograph (b) the two sweeps are displaced in phase thereby permitting an examination of the individual traces; oscillograph (c) shows three adjacent repeller modes, again with the forward and reverse sweeps superimposed. This absence of multiple transit hysteresis is, perhaps, surprising in view of the reflector space geometry of Fig. 16. From this figure we might expect a considerable fraction of the returning electrons to re-enter the cathode region. This has, indeed, been verified in an experiment (see Fig. 22) in which cathode current was monitored as the repeller voltage was varied over a wide range. Because of the limited power handling capacity of the repeller, this experiment was performed with a pulsed beam. For values of repeller voltage negative with respect to the cathode, the cathode current and hence the apparent perveance was about 25 per cent lower than that corresponding to positive repeller voltages. The absence of hysteresis must therefore mean either that few electrons reach the interaction gap for a third transit or that those which do are no longer bunched.

ACKNOWLEDGMENTS

The M1805 millimeter wave reflex klystron is the outcome of an intensive team effort which has required and received the cooperation of numerous persons working in highly diversified fields. In presenting this paper, therefore, the author wishes to be considered as the spokesman for this group. Of the people most closely associated with this project from its inception, L. B. Luckner has been in large measure responsible for the mechanical design and its reduction to practice, while F. P. Drechsler has contributed importantly to other phases of this work, notably the design of the electron gun, the fabrication of the ribbon grid and the electrical testing. The feasibility of hubbing as applied to the cavity block was first established by L. B. Luckner working in close cooperation with H. O. Schroder, subsequent refinements and simplifications being due to W. Patterson and K. E. Schukraft of the Precision Room at Murray Hill. The group of skilled technicians responsible for the assembly of the M1805 was under the direction of H. W. Schwarz. Other contributors to this project — but by no means all — have been named in the body of the text. In addition, the success of this project is in no small measure due to the encouragement received from J. O. McNally, J. R. Wilson and H. T. Friis. Finally, the author would like to thank D. A. Chisholm for his constructive criticism of the manuscript.

REFERENCES

1. S. E. Miller, Waveguide as a Communication Medium, B. S. T. J., Nov., 1954.
2. J. R. Pierce and W. G. Shepherd, Reflex Oscillators, B. S. T. J., pp. 603–607, July, 1947.
3. O. Heil and J. J. Ebers, A New Wide-Range High Frequency Oscillator, Proc. I.R.E., June, 1950.
4. J. R. Pierce and W. G. Shepherd, Reflex Oscillators, B. S. T. J., pp. 478, July, 1947.
5. J. S. Thorp and D. J. Taylor, RF Conductivity in Copper at 8 mm Wavelengths, Services Electronics Research Laboratory Technical Journal (England), **3,** No. 2, July, 1953.
6. J. S. Thorp, RF Conductivity in Copper at 8-mm Wavelengths, Proc. I.E.E., (England) **101,** Part III, p. 357, Nov., 1954.
7. U. S. Patent No. 2,296,885 by R. L. Vance, 1942.

Interchannel Interference in FM and PM Systems Under Noise Loading Conditions

By W. R. BENNETT, H. E. CURTIS, and S. O. RICE

(Manuscript received January 5, 1955)

One of the principal sources of interchannel interference in multichannel FM and PM systems is the dependence of the attenuation and phase shift of the transmission path on frequency. Here we study the interference produced by a simple kind of such a dependence, namely that due to a single echo. Besides being important in themselves, the results are of interest because they may be used to estimate the interference in other, more complicated cases.

In this paper expressions are derived for the interchannel interference produced by echoes of relatively small amplitude. Several important special cases are studied in detail and curves that simplify the computation of the interference power are presented.

CONTENTS

Section	*Page*
Introduction	601
1. Development of General Formulas	602
2. Applications to Flat Noise Signal	609
Exact Expression, Equations (2.3)–(2.8)	610
Ratio of Interference to Signal — PM	611
Ratio of Interference to Signal — FM	611
3. Approximations — Phase Modulation	611
Table 3.1	612
4. Approximations — Frequency Modulation	614
Table 4.1	615
5. Interchannel Interference Power	620
6. Use of Equivalent Echo to Estimate Interchannel Interference	628

INTRODUCTION

The systems we shall consider are of the FDM-FM and FDM-PM types; that is, systems in which the composite signal wave (the "baseband signal") from a group of carrier telephone channels in frequency division multiplex (FDM) is transmitted by frequency modulation (FM) or phase modulation (PM). Such methods are currently being used in

the Bell System to send large groups of telephone channels by microwave radio. If the FM signal is accompanied by echoes, which may be due to reflections in the equipment or in the transmission medium, the wave of instantaneous frequency versus time is distorted in a nonlinear manner and interchannel interference occurs. Here we shall be concerned with this interference.

The distortion has been analyzed in a number of previous publications[1-4] for the case in which the base-band signal may be represented by one or more sine waves. However, when the number of telephone channels is not small, the sine wave representation becomes unwieldy because a large number of both low and high order modulation products must be considered. Here we avoid this difficulty, at the cost of somewhat more complex analysis, by using a band of random noise to represent the multiplex signal.

It has been found in practice that such a random noise signal of appropriate bandwidth and power adequately simulates a composite speech signal. For studies involving interchannel interference the energy corresponding to some particular telephone channel is removed. When such a wave is impressed on the frequency modulator and the resulting FM wave is transmitted, detected, and finally demodulated, the received output in the originally clear channel represents interchannel interference. Measurements of this type have been discussed by Albersheim and Schafer.[3]

One of the principal sources of interchannel interference is the variation of the attenuation and phase shift of the transmission path with frequency. An echo produces a simple form of such a variation. In fact, it is often possible to estimate the effect produced by a more complex type of variation by comparing it with a roughly equivalent echo.

It is the purpose of this paper to develop formulas for the interchannel interference produced by echoes of relatively small amplitude in FM and PM systems. General expressions are given in the form of definite integrals which may be evaluated by numerical integration. Approximations are obtained in a number of cases of importance and numerical tables and curves are furnished to facilitate applications to specific problems.

We wish to express our thanks to Miss Mary Corr and Miss Barbara Fischer who have performed the rather lengthy computations required for this investigation.

1. DEVELOPMENT OF GENERAL FORMULAS

Since the tolerable amount of interchannel interference in multiplex telephony is small, the practically important situations are characterized

by echoes which are relatively small relative to the main transmitted wave. It is necessary to consider in detail only the case of one small echo. Effects of small multiple echoes may then be calculated by superposition of the effects of single echoes. Our distortion problem then reduces to the following: An original signal

$$E_0(t) = E \sin [pt + \varphi(t)] \tag{1.1}$$

is received along with an echo

$$E_1(t) = rE \sin [p(t - T) + \varphi(t - T)] \tag{1.2}$$

Here r represents the ratio of echo amplitude to the amplitude of the principal received component, p is the unmodulated carrier frequency, T is the delay difference of the paths and $\varphi(t)$ is the phase variation caused by the multichannel signal.

Calculation of the phase of the original signal plus its echo is equivalent to calculating the value of θ in the representation

$$\sin x + r \sin (x + y) = V \sin (x + \theta) \tag{1.3}$$

with x, y, r, V, and θ real numbers. V and θ are determined as functions of r and y by the two equations obtained from the sin x and cos x portions of (1.3). When $r \ll 1$, V approaches unity, and θ is proportional to r; direct expansion in powers of θ shows that to a first order approximation

$$\theta = r \sin y \tag{1.4}$$

In our case,

$$x = pt + \varphi(t) \tag{1.5}$$

$$y = \varphi(t - T) - \varphi(t) - pT \tag{1.6}$$

Hence the phase error produced by the echo is

$$\theta(t) = r \sin [v(t) - pT], \tag{1.7}$$

$$v(t) = \varphi(t - T) - \varphi(t). \tag{1.8}$$

Our problem is to study the power spectrum of $\theta(t)$ when $\varphi(t)$ is a random noise wave.

The treatment will be restricted to the practically important case in which the noise source is free from dc and discrete sinusoidal components. A sufficient mathematical condition, which we shall adopt to insure this, is that the power spectrum of the noise have limited total fluctuation. We shall assume also that the noise wave is a member of a Gaussian ensemble and hence that all the statistical properties are derivable from

either its power spectrum or autocorrelation function. Which of the two quantities we use at any stage of the computations is a matter of relative expediency. We shall adopt a uniform notation in that an ensemble of time functions $x(t)$ will be said to have a power spectrum $w_x(f)$ and an autocorrelation function $R_x(\tau)$, i.e.,

$$R_x(\tau) = \text{ave } [x(t)x(t+\tau)] = \int_0^\infty w_x(f) \cos 2\pi\tau f \, df \tag{1.9}$$

$$w_x(f) = 4\int_0^\infty R_x(\tau) \cos 2\pi f\tau \, d\tau \tag{1.10}$$

The power spectrum is proportional to the mean square of the response of a filter of bandwidth df centered at f when $x(t)$ is applied as input.

In the echo distortion problem, the power spectrum $w_\varphi(f)$ is the quantity given and the power spectrum $w_\theta(f)$ is the one desired. The corresponding autocorrelation functions $R_\varphi(\tau)$ and $R_\theta(\tau)$ are calculable from the corresponding power spectra and vice versa. The power spectrum is the more convenient choice as a function to compute when we deal with the noise response of a linear system with known transfer admittance $Y(if)$, since the power spectrum of the response is merely $|\, Y(if)\,|^2$ times that of the input. For example, differentiation is equivalent to transmission through a transfer admittance $i2\pi f$ and hence multiplies the power spectrum by $4\pi^2 f^2$. The inverse process of integration divides the power spectrum by $4\pi^2 f^2$. When a highly nonlinear operation is performed on the noise, or when the operation is a linear one more simply described in the time domain, the autocorrelation function may be simpler to compute.

The first step in the solution of the problem is to evaluate the statistics of the noise wave ensemble,

$$\{v(t)\} = \{\varphi(t - T) - \varphi(t)\} \tag{1.8}$$

in terms of the statistics of $\varphi(t)$. This can readily be done in terms of power spectra by calculating the transfer admittance of a two-path transmission system such as the right-hand member of (1.8) defines, or in terms of autocorrelation functions by averaging the appropriate time functions. The latter procedure gives:

$$\begin{aligned} R_v(\tau) &= \text{ave } [v(t)\, v(t+\tau)] \\ &= \text{ave } \{[\varphi(t - T) - \varphi(t)][\varphi(t - T + \tau) - \varphi(t + \tau)]\} \\ &= \text{ave } [\varphi(t - T)\, \varphi(t - T + \tau)] + \text{ave } [\varphi(t)\, \varphi(t + \tau)] \\ &\qquad - \text{ave } [\varphi(t - T)\, \varphi(t - T + \tau + T)] \\ &\qquad - \text{ave } [\varphi(t)\, \varphi(t + \tau - T)] \\ &= 2R_\varphi(\tau) - R_\varphi(\tau + T) - R_\varphi(\tau - T) \end{aligned} \tag{1.11}$$

In the averaging process we have made use of the facts that the average of the sum is equal to the sum of the averages and that $\varphi(t)$ is a stationary ensemble, allowing us to replace $t - T$ by t when averaging. Since the two-path transmission system is linear and invariable, the $v(t)$ ensemble remains Gaussian. The corresponding power spectra relations are, from (1.10),

$$w_v(f) = 4\, w_\varphi(f) \sin^2 \pi f T \tag{1.12}$$

Our next step called for by (1.7) requires the evaluation of the autocorrelation function of a sine function of a band of noise. The solution may be obtained* from (3.2 — 7) of Reference 5, which gives a general theorem for a Gaussian noise ensemble expressible in our notation as

$$\text{ave } \{\exp [iax(t) + ibx(t + \tau)]\} = \exp \left[-\frac{a^2 + b^2}{2} R_x(0) - abR_x(\tau)\right] \tag{1.13}$$

In this equation a and b are real constants. Multiplying both sides by the constant $\exp(i\beta)$, β real, and equating real parts gives the more directly applicable result:

$$\text{ave } \{\cos [ax(t) + bx(t + \tau) + \beta]\} = \exp\left[-\frac{a^2 + b^2}{2} R_x(0) - abR_x(\tau)\right] \cos \beta \tag{1.14}$$

Referring now to (1.7) we write

$$\begin{aligned} R_\theta(\tau) &= \text{ave } \{r \sin [v(t) - pT]\, r \sin [v(t + \tau) - pT]\} \\ &= \frac{r^2}{2} \text{ave } \{\cos [v(t) - v(t + \tau)]\} \\ &\qquad - \frac{r^2}{2} \text{ave } \{\cos [v(t) + v(t + \tau) - 2pT]\} \qquad (1.15) \\ &= \frac{r^2}{2} \exp [- R_v(0) + R_v(\tau)] \\ &\qquad - \frac{r^2}{2} \exp [- R_v(0) - R_v(\tau)] \cos 2pT \\ &= \frac{r^2}{2} e^{-R_v(0)} [e^{R_v(\tau)} - e^{-R_v(\tau)} \cos 2pT] \end{aligned}$$

* This application of the characteristic functions method of attack has been employed in similar situations by D. Middleton in Reference 6 and M. K. Zinn in unpublished memoranda.

Note that since the operation of taking the sine is nonlinear, the $\theta(t)$ ensemble is not Gaussian and neither the power spectrum nor the autocorrelation is sufficient to give a complete statistical description. Either will be sufficient for our purposes however.

The power spectrum of the distortion ensemble $\theta(t)$ is now determinable from the Fourier cosine transform of 4 $R_\theta(\tau)$ as indicated by (1.10). The remaining steps are simplified, however, if we perform two preliminary operations on $R_\theta(\tau)$ before the final Fourier transform is calculated.

We first observe that if $R_\theta(\tau)$ contains a component C which does not vary with τ, the ensemble $\theta(t)$ contains a dc component $\bar{\theta} = \sqrt{C}$. The presence of C complicates the integration and hence we subtract out such a term before taking the Fourier transform. The value of C is given by

$$\begin{aligned} C &= \lim_{\tau\to\infty} R_\theta(\tau) \\ &= \frac{r^2}{2} e^{-R_v(0)} \lim_{\tau\to\infty} [e^{R_v(\tau)} - e^{-R_v(\tau)} \cos 2pT] \\ &= \frac{r^2}{2} e^{-R_v(0)}(1 - \cos 2pT) \end{aligned} \tag{1.16}$$

In calculating the limit we have made use of the fact that, for our assumed noise wave, $R_v(\tau)$ must approach zero as τ becomes infinite. Our autocorrelation function of interest is thereby reduced to

$$R_{\theta-\bar{\theta}}(\tau) = \frac{r^2}{2} e^{-R_v(0)} \{e^{R_v(\tau)} - 1 - [e^{-R_v(\tau)} - 1] \cos 2pT\} \tag{1.17}$$

which is the autocorrelation function of the ensemble $\theta(t) - \bar{\theta}$.

Our second preliminary operation on $R_\theta(\tau)$ is suggested by our ultimate goal, namely the calculation of the interchannel interference spectrum $w_c(f)$ (i.e., $w_c(f)\,df$ is the average power received in an idle channel of width df when the other channels are carrying signals). We note that the spreading of the original spectrum into initially vacant frequency ranges occurs solely because of the nonlinear dependence of $R_{\theta-\bar{\theta}}(\tau)$ on $R_v(\tau)$. The disturbance received in an idle channel is therefore produced by this nonlinearity. The part of $R_{\theta-\bar{\theta}}(\tau)$ which varies linearly with $R_v(\tau)$ may be expected to represent the linear transmission, and the difference between $R_{\theta-\bar{\theta}}(\tau)$ and its linear portion to represent the nonlinear transmission. In other words, subtracting the linear portion from $R_{\theta-\bar{\theta}}(\tau)$ is equivalent to removing the linear transmission from any channel without disturbing the nonlinear contributions from the re-

maining channels. These considerations lead us to set the autocorrelation function $R_c(\tau)$ corresponding to $w_c(f)$ equal to $R_{\theta-\bar{\theta}}(\tau)$ minus its linear portion. The work of Appendix I shows that this equality is rigorously true.

In order to perform the subtraction it is convenient to write (1.17) in the form

$$R_{\theta-\bar{\theta}}(\tau) = F[R_v(\tau)] \tag{1.18}$$

where in our case for a general variable z,

$$F(z) = \frac{r^2}{2} e^{-R_v(0)}[e^z - 1 - (e^{-z} - 1)\cos 2pT] \tag{1.19}$$

The portion of $R_{\theta-\bar{\theta}}(\tau)$ which varies linearly with $R_v(\tau)$ is $F'(0)\ R_v(\tau)$ where

$$F'(z) = \frac{r^2}{2} e^{-R_v(0)}(e^z + e^{-z}\cos 2pT) \tag{1.20}$$

$$F'(0) = \frac{r^2}{2} e^{-R_v(0)}(1 + \cos 2pT) \tag{1.21}$$

According to the foregoing discussion and Appendix I, the autocorrelation function and the power spectrum of the interchannel interference are given by

$$\begin{aligned} R_c(\tau) &= R_{\theta-\bar{\theta}}(\tau) - F'(0)R_v(\tau) \\ &= \frac{r^2}{2} e^{-R_v(0)}\{e^{R_v(\tau)} - 1 - R_v(\tau) \\ &\qquad - [e^{-R_v(\tau)} - 1 + R_v(\tau)]\cos 2pT\} \end{aligned} \tag{1.22}$$

$$w_c(f) = 4\int_0^\infty R_c(\tau)\cos 2\pi f\tau\, d\tau \tag{1.23}$$

The power spectrum of $\theta(t) - \bar{\theta}$ may be simply expressed in terms of $w_c(f)$ without further integration. We find by applying the Fourier transform relationship of (1.9) and (1.10) to (1.22):

$$\begin{aligned} w_{\theta-\bar{\theta}}(f) &= 4\int_0^\infty [R_c(\tau) + F'(0)R_v(\tau)]\cos 2\pi f\tau\, d\tau \\ &= w_c(f) + F'(0)w_v(f) \\ &= w_c(f) + 4F'(0)w_\varphi(f)\sin^2 \pi fT \\ &= w_c(f) + 2r^2e^{-R_v(0)}(1 + \cos 2pT)w_\varphi(f)\sin^2 \pi fT \end{aligned} \tag{1.24}$$

where we have used (1.12) and (1.21). The power spectrum $w_\theta(f)$ of θ is obtained by adding to (1.24) a spire at $f = 0$ to represent the power in $\bar{\theta}$, the dc portion of θ.

The results obtained up to this point may be summarized as follows: Let $\varphi(t)$ be the phase variation produced by the impressed multichannel signal. The echo (1.2) produces a phase error $\theta(t)$ (1.7) in the received signal. $\theta(t)$ has a dc component $\bar{\theta}$ given by the square root of (1.16). $\theta(t) - \bar{\theta}$ is a function of time which fluctuates about zero and has the power spectrum $w_{\theta-\bar{\theta}}(f)$ given by (1.24). Now consider the problem of computing the interchannel interference. One procedure would be to consider the case in which all but one of the channels are loaded. Then the power spectrum $w_\varphi(f)$ of $\varphi(t)$ would have a narrow slot in it corresponding to the zero power in the unloaded channel. The values of $w_{\theta-\bar{\theta}}(f)$ computed from this $w_\varphi(f)$ for values of f within the slot would give the channel interference spectrum (for phase modulation). However, it turns out that as the slot width approaches zero, these values of $w_{\theta-\bar{\theta}}(f)$ approach those of $w_c(f)$ where $w_c(f)$ is computed from (1.23) on the assumption that all channels are loaded. In other words the $w_\varphi(f)$ used in the calculation of (1.23) has no slot. (Actually the same value of $w_c(f)$ is obtained whether $w_\varphi(f)$ has a slot or not. Of course, the slot must be vanishingly narrow).

Almost all of the preceding work pertains to the power spectrum of the phase error $\theta(t)$. For the sake of completeness we shall give the power spectrum of the complete phase angle, $Q(t) = \varphi(t) + \theta(t)$, of the output. In order to obtain all of the $0(r^2)$ term it is necessary to add another term to the approximation (1.7):

$$\theta(t) = r \sin [v(t) - pT] - \frac{r^2}{2} \sin [2v(t) - 2pT] + 0(r^3) \quad (1.25)$$

The autocorrelation function for $Q(t)$ may be shown to be

$$\begin{aligned} R_Q(\tau) &= R_\varphi(\tau) + R_\theta(\tau) + \text{ave } [\varphi(t)\,\theta(t+\tau) + \varphi(t+\tau)\,\theta(t)] \\ &= R_\varphi(\tau) + R_\theta(\tau) - rR_v(\tau)\, e^{-R_v(0)/2} \cos pT \\ &\quad + r^2 R_v(\tau)\, e^{-2R_v(0)} \cos 2pT + 0(r^3) \end{aligned} \quad (1.26)$$

The average has been evaluated by using

$$\begin{aligned} &\text{ave } \{\varphi(t \pm \tau) \sin [\varphi(t - T) - \varphi(t) - pT]\} \\ &\quad = [R_\varphi(T \pm \tau) - R_\varphi(\tau)] \cos pT \exp [R_\varphi(T) - R_\varphi(0)] \end{aligned} \quad (1.27)$$

and the corresponding result for 2φ and $2pT$ ($R_{2\varphi}(\tau) = 4R_\varphi(\tau)$). Equation (1.27) may be obtained from the result similar to (1.13) which con-

tains the three variables $\varphi(t)$, $\varphi(t \pm \tau)$, $\varphi(t - T)$ instead of $x(t)$ and $x(t + \tau)$.

The power spectrum $w_Q(f)$ of $Q(t)$ may be obtained from (1.26) by replacing the autocorrelation functions in the coefficients by the corresponding power spectra. By using (1.24) we may obtain the following expression for the power spectrum of the fluctuating portion $Q(t) - \bar{Q}$ of $Q(t)$, $\bar{Q} = \bar{\theta}$ being the dc portion of $Q(t)$:

$$w_{Q-\bar{Q}}(f) = w_\varphi(f) + w_c(f) + w_v(f)[-r\, e^{-R_v(0)/2} \cos pT + r^2\, e^{-R_v(0)} \cos^2 pT + r^2\, e^{-2R_v(0)} \cos 2pT] \qquad (1.28)$$

The expression (1.12) for $w_v(f)$ shows that it is proportional to $w_\varphi(f)$. Hence, assuming $w_\varphi(f)$ to have a narrow slot corresponding to an idle channel, it is seen that (1.28) is in agreement with the fact that only $w_c(f)$ contributes to the interchannel interference spectrum.

We remark that for echoes produced by reflections in wave guides the angle $2pT$ in (1.22) and (1.24) usually contains many multiples of 2π. In most cases it would in fact be reasonable to average the effect of the term $\cos 2pT$ by assuming the angle to be uniformly distributed throughout 2π radians. This gives the result zero for the term since plus and minus values are symmetrically distributed. We shall, however, carry the term along in our formulas since it is of importance when the delays are small, as in multipath radio propagation, and when our results are used to estimate interchannel interference in general.

2. APPLICATIONS TO FLAT NOISE SIGNAL

In general $w_\varphi(f)$ depends on the type of preemphasis used in the channel multiplex signal. Two representative conditions will be studied: (1) phase modulation (PM) in which channels at equal level are impressed on a phase modulator, or the more usual equivalent situation of equal level channels which are differentiated before being impressed on a frequency modulator, and (2) frequency modulation (FM) in which channels at equal level are impressed on a frequency modulator. The appropriate power spectra are:

$$\text{PM:} \quad w_\varphi(f) = P_0, \qquad f_a < f < f_b \qquad (2.1)$$

$$\text{FM:} \quad w_\varphi(f) = P_0/4\pi^2 f^2, \qquad f_a < f < f_b \qquad (2.2)$$

The latter form results because the phase is the time integral of the instantaneous frequency, which has a flat spectrum. In PM, P_0 is expressed in $(\text{radians})^2/\text{cps}$, but in FM, P_0 is expressed in $(\text{radians/sec.})^2/\text{cps}$. In

the case of most common practical interest, the band of frequencies between 0 and f_a is relatively narrow compared to the range $f_b - f_a$. We have accordingly assumed that f_a approaches zero. In the FM case we cannot set $f_a = 0$ immediately because the power spectrum would then become unbounded at the origin. It is found, however, that finite limits are approached for the actual quantities of interest which we compute.

When the spectrum $w_\varphi(f)$ of the signal has the form (2.1) or (2.2), expression (1.23) for the interchannel interference spectrum may be written as

$$w_c(f) = r^2(2\pi f_b)^{-1}(G - H \cos 2pT) \tag{2.3}$$

where $w_c(f)\, df$ is measured in (radians)^2 and

$$G = 2e^{-R_v(0)} \int_0^\infty [e^{R_v(\tau)} - R_v(\tau) - 1] \cos au\, du \tag{2.4}$$

$$H = 2e^{-R_v(0)} \int_0^\infty [e^{-R_v(\tau)} + R_v(\tau) - 1] \cos au\, du \tag{2.5}$$

$$u = 2\pi f_b \tau \qquad U = 2\pi f_b T \qquad a = f/f_b \tag{2.6}$$

For PM:

$$R_v(\tau) = P_0 f_b \left[\frac{2 \sin u}{u} - \frac{\sin (u + U)}{u + U} - \frac{\sin (u - U)}{u - U}\right] \tag{2.7}$$

where $P_0 f_b$ is the mean square value, in (radians)^2, of the PM signal $\varphi(t)$.

For FM:

$$R_v(\tau) = A[-2(1 - \cos U)\cos u - 2uSi(u) + (u + U)Si(u + U) + (u - U)Si(u - U)] \tag{2.8}$$

$$A = (P_0 f_b)/(2\pi f_b)^2$$

In (2.8) $P_0 f_b$ is the mean square value (measured in (radians per second)^2) of the FM signal $\varphi'(t)$. Consequently, $A = (\sigma/f_b)^2$ where σ is the rms frequency deviation, in cycles per second, of the signal. Equations (2.7) and (2.8) follow from (2.1), (2.2) and (1.11).

When $R_v(0)$ is small compared to unity, the same is true of $R_v(\tau)$, and (2.4) and (2.5) lead to the approximations

$$G \approx H \approx \int_0^\infty R_v^2(\tau) \cos au\, du \tag{2.9}$$

The interchannel interference calculated from (2.9) is that produced by

the "second order modulation products" and is studied, together with other approximations, in Sections 3 and 4.

The quantity of interest in practice is the ratio of average interference power in the idle channel to average signal power in an adjacent channel. For PM the average interference and signal powers (in a narrow channel centered on frequency f) are the channel bandwidth times $w_c(f)$ and $w_\varphi(f)$, respectively. When the number of channels is large, the power spectrum does not change appreciably in going from one channel to the next and we may write

$$\frac{P_I}{P_S} = \frac{w_c(f)}{w_\varphi(f)} = \frac{r^2}{2\pi P_0 f_b}(G - H\cos 2pT) \tag{2.10}$$

as the desired interference-to-signal power ratio.

For FM the average interference and signal powers in a narrow channel are $(2\pi f)^2 \times$ (channel bandwidth) times $w_c(f)$ and $w_\varphi(f)$, respectively. These powers are measured in (radians per second)^2. When we take the ratio P_I/P_S, the $(2\pi f)^2 \times$ (channel bandwidth) cancels out and we have from (2.2) and (2.3)

$$\frac{P_I}{P_S} = \frac{r^2a^2}{2\pi A}(G - H\cos 2pT) \tag{2.11}$$

3. APPROXIMATIONS FOR G AND H — PHASE MODULATION

Table 3.1 contains various approximate expressions for the quantities G and H which enter expression (2.3) for the channel interference spectrum $w_c(f)$. The notation is explained in Section 2, and the results apply to the case in which $w_\varphi(f)$ has the flat power spectrum (2.1).

Case 1 gives the exact expressions developed in Section 2. Case 2 is the "second order modulation approximation", valid when $R_v(0) \ll 1$. Evaluation of the integral (2.9) for G with the help of

$$\int_{-\infty}^{\infty} \frac{\sin u \sin (u - U)}{u(u - U)} \cos au\, du$$

$$= \begin{cases} \pi U^{-1} \cos\left(\dfrac{aU}{2}\right) \sin\left(U - \dfrac{aU}{2}\right), & 0 \leqq a \leqq 2 \\ 0, & 2 \leqq a \end{cases}$$

leads to the following expression for the quantity J appearing in Table 3.1:

TABLE 3.1 — PHASE MODULATION

Case No.	Restrictions on Parameters U and P_0f_b	G and H	Notes
1	No restrictions, $w_\varphi(f)$ defined by (2.1)	G defined by (2.4) H defined by (2.5) $R_v(\tau)$ defined by (2.7)	P_0f_b = ave $[\varphi^2(t)]$ $\frac{P_I}{P_s} = \frac{r^2}{2\pi P_0 f_b}$ $(G - H\cos 2pT)$, $U = 2\pi f_b T$, $a = f/f_b$
2	$2P_0f_b\left(1 - \frac{\sin U}{U}\right) \ll 1$	$G \approx H \approx 2\pi(P_0f_b)^2J$, J defined by (3.1)	"2nd Order Modulation" approx., J tabulated in Table 3.2, $10\log_{10} J$ plotted in Fig. 5.1
3	$P_0f_bU^2/3 \ll 1$, $U \ll 1$	$G \approx H \approx 2\pi(P_0f_b)^2\frac{U^4}{240}$ $\cdot[12 - 30a + 20a^2 - a^5]$	Special case of Case 2
4	$2P_0f_b \ll 1$, $U \gg 1$	$G \approx H \approx 2\pi(P_0f_b)^2$ $\cdot\left(1 - \frac{a}{2}\right)\left(1 + \frac{\cos aU}{2}\right)$	Special case of Case 2 G is a rapidly oscillating function of a
5	$P_0f_bU^2/3 \gg 1$, $U \ll 1$	$G \gg H$, $G \approx (10\pi/P_0f_bU^2)^{1/2}$ $\exp(-10a^2/4P_0f_bU^2)$	When $U \ll 1$ Case 3 applies when P_0f_b is small Case 5 applies when P_0f_b is large
6	$U \gg 1$	$G \approx e^{-b_0}[I(b_0, a) + 2I(b_1, a)\cos aU]$, $H \approx e^{-b_0}[I(-b_0, a) + 2I(-b_1, a)\cos aU]$, $b_0 = 2P_0f_b$, $b_1 = -P_0f_b$ $I(b, a)$ studied in Appendix III	When $P_0f_b \ll 1$, Case 6 reduces to Case 4. When $P_0f_b \gg 1$, $G \gg H$. As a increases, G oscillates between G^+ and G^- given in Table 3.3. See Table 5.1

$$J = \left(1 - \frac{a}{2}\right)\left(1 + \frac{\cos aU}{2}\right) + \frac{\sin(2U - aU)}{4U} - \frac{2}{U}\cos\left(\frac{aU}{2}\right)\sin\left(U - \frac{aU}{2}\right) \tag{3.1}$$

Values of J are tabulated in Table 3.2 for various values of a and k where $U = k\pi/4$ (or $T = k/(8f_b)$). J is zero when a exceeds 2.

Cases 3 and 4 in Table 3.1 follow directly from Case 2. In order to

TABLE 3.2 — VALUES OF J

k	$a = 0$	0.25	0.50	0.75	1.00	1.25
0	0	0	0	0	0	0
1	0.018	0.009	0.003	0.001	0.002	0.004
2	0.227	0.115	0.041	0.011	0.023	0.057
3	0.794	0.434	0.161	0.049	0.098	0.231
4	1.500	0.903	0.352	0.123	0.250	0.524
5	1.924	1.284	0.527	0.224	0.458	0.816
6	1.924	1.385	0.585	0.332	0.659	0.936
7	1.712	1.231	0.505	0.427	0.773	0.800
8	1.500	0.994	0.375	0.506	0.750	0.494
9	1.335	0.800	0.325	0.579	0.602	0.228
10	1.245	0.658	0.425	0.657	0.405	0.180
11	1.307	0.544	0.643	0.725	0.262	0.358
12	1.500	0.472	0.883	0.752	0.250	0.601

obtain Case 5 we note that when $U \ll 1$, expression (2.7) gives

$$R_v(\tau) \approx -U^2 P_0 f_b \frac{d^2}{du^2} \frac{\sin u}{u}. \tag{3.2}$$

The corresponding integrals for G and H could, if required, be investigated by the methods used to study Lewin's integral in Appendix III. However here we consider only the case where $U^2 P_0 f_b$ is so large that (1) $\exp R_v(\tau)$ is the dominant term in the integral (2.4) for G, and (2) most of the contribution to the value of the integral comes from the region around $u = 0$. The results for Case 5 then follow from (2.4) and the fact that (3.2) becomes

$$R_v(\tau) \approx U^2 P_0 f_b \left(\frac{1}{3} - \frac{u^2}{10}\right)$$

When $U \gg 1$, expression (2.7) shows that $R_v(\tau)$ is small except when u is near 0 or near U:

$$\begin{aligned} R_v(\tau) &\approx 2\, P_0 f_b\, u^{-1} \sin u, && u \text{ near } 0 \\ R_v(\tau) &\approx -P_0 f_b (u - U)^{-1} \sin (u - U), && u \text{ near } U \end{aligned} \tag{3.3}$$

These approximations, expression (2.4) for G, and the definition (A3–1) of $I(b, a)$ give the results stated in Case 6. Just as in Case 4, G is a rapidly oscillating function of a when U is large. It oscillates between G^+ and G^- where

$$G^{\pm} = e^{-b_0}[I(b_0, a) \pm 2I(b_1, a)] \tag{3.4}$$

Values of G^+ and G^- are given in Table 3.3 for various values of $a = J/f_b$ and $P_0 f_b$ [in (radians)^2].

TABLE 3.3 — VALUES OF G^+ AND G^-. THE UPPER NUMBER OF AN ENTRY IS G^+ AND THE LOWER NUMBER IS G^-.

$P_0 f_b$ (radians)^2	$a = 0$	0.25	0.50	0.75	1.00	1.25
0.25	0.384 0.160	0.338 0.144	0.292 0.126	0.245 0.107	0.197 0.087	0.148 0.066
0.50	1.018 0.504	0.906 0.464	0.789 0.415	0.665 0.357	0.537 0.291	0.406 0.222
0.75	1.55 0.89	1.39 0.83	1.22 0.75	1.04 0.65	0.845 0.533	0.645 0.411
1.00	1.90 1.22	1.73 1.15	1.53 1.05	1.32 0.92	1.07 0.76	0.830 0.596
2.00	2.16 1.86	2.04 1.80	1.88 1.68	1.68 1.52	1.43 1.31	1.17 1.07
4.00	1.59 1.57	1.56 1.54	1.50 1.48	1.40 1.40	1.28 1.28	1.14 1.14

4. APPROXIMATIONS FOR G AND H — FREQUENCY MODULATION

The various cases which we shall consider for FM are roughly similar to those considered in Section 3 for PM, and are listed in Table 4.1. The power spectrum $w_\varphi(f)$ is assumed to be that given by (2.2). As pointed out in Section 2, the average FM signal power in a narrow channel of width Δf centered on frequency f is assumed to be $P_0\Delta f$ (radians/second)^2 if $0 < f < f_b$ and zero if $f_b < f$. The average interchannel interference power is $(2\pi f)^2\, w_c(f)\Delta f$ (radians/second)^2. When $f > f_b$ this gives all of the power present in the frequency interval Δf.

Case 1, Table 4.1, gives the exact expressions for G and H, and Case 2 corresponds to the "second order modulation" approximation which holds when $R_v(0) \ll 1$, $R_v(0)$ being computed from (2.8). K is a function of $a = f/f_b$ and $U = 2\pi f_b T$ which may be obtained by writing the integral (2.9) as

$$G \approx \pi f_b \int_{-\infty}^{\infty} R_v^{\,2}(\tau) e^{2\pi i f\tau}\, d\tau = 4^{-1}\pi f_b \int_{-\infty}^{\infty} w_v(x) w_v(f - x)\, dx \quad (4.1)$$

where $R_v(\tau)$ and $w_v(f)$ are taken to be even functions. From the definitions (1.12) and (2.2) for $w_v(f)$ and $w_\varphi(f)$ we have

$$w_v(f) = \begin{cases} P_0(\pi f)^{-2} \sin^2 \pi f T, & |f| < f_b \\ 0, & |f| > f_b \end{cases} \quad (4.2)$$

TABLE 4.1 — FREQUENCY MODULATION

Case No.	Restrictions on Parameters U and A	G and H	Notes
1	No restriction, $w_\varphi(f)$ defined by (2.2)	G defined by (2.4) H defined by (2.5) $R_v(\tau)$ defined by (2.8)	$A = (\sigma/f_b)^2$, σ = rms $[d\varphi(t)/dt]/2\pi$, σ = rms frequency deviation of signal measured in cycles/sec. $P_I/P_S = [r^2a^2/(2\pi A)] \cdot [G - H \cos 2pT]$
2	$2A[USi(U) - 1 + \cos U] \ll 1$	$G \approx H \approx 2\pi A^2a^{-2}UK$ K defined by (4.5) and (4.6)	"2nd Order Modulation" approx. K tabulated in Table 4.2. See Fig. 5.2
3	$AU^2 \ll 1$, $U \ll 1$.	$G \approx H \approx \pi A^2U^4(2 - a)/4$, $0 \leqq a \leqq 2$	Special case of Case 2 $UK \approx a^2U^4(2 - a)/8$
4	$AU\pi \ll 1$, $U \gg 1$	$G \approx H \approx \begin{cases} 2\pi^2A^2Ua^{-2}, & 0 < a < 1 \\ \pi^2A^2U, & a = 1 \\ 0, & a > 1 \end{cases}$	Special case of Case 2
5	$U \ll 1$, Lewin's case	$G \approx e^{-b}I(b, a)$ $H \approx e^{-b}I(-b, a)$ $b = AU^2$	Case 5 agrees with Case 3 when $b \ll 1$. $G \gg H$ for $b \gg 1$. $I(b, a)$ defined and tabulated in Appendix III. See Fig. 5.4.
6	$4A \gg 1$, $U \gg 1$	$G \approx [\pi y_1/(A \sinh y_1)]^{1/2} \exp[-2A(\cosh y_1 - 1)]$ y_1 defined by $\frac{a}{2A} = \int_0^{y_1} \frac{\sinh v}{v}\, dv$	$G \gg H$ Value of G is independent of U. See Fig. 5.3.
7	$B \gg 1$, $U \approx 1$, $B = A[1 - U^{-1}\sin U]$	$G \approx [\pi/B]^{1/2} \exp[-a^2/(4B)]$	$G \gg H$

where the lower limit f_a of the frequency band is taken to be very close to zero. When $f > 2f_b$ the value of (4.1) is zero. When we take $0 < f < 2f_b$ the limits of integration in (4.1) are $x = f - f_b$ and $x = f_b$. Changing the variable of integration in (4.1) from x to $y = 2\pi xT$ converts (4.1) into

$$4\pi A^2U^3 \int_{\alpha-U}^{U} y^{-2}(\alpha - y)^{-2} \sin^2 \frac{y}{2} \sin^2 \left(\frac{\alpha - y}{2}\right) dy \tag{4.3}$$

where $\alpha = 2\pi fT = aU$. By partial fractions

$$\alpha^2 y^{-2}(\alpha - y)^{-2} = y^{-2} + 2\alpha^{-1}y^{-1} + (\alpha - y)^{-2} + 2\alpha^{-1}(\alpha - y)^{-1} \tag{4.4}$$

Considerations of symmetry show that the $\alpha - y$ terms on the right contribute the same amount to (4.3) as do the y terms. When the y^{-2} term is converted into y^{-1} by an integration by parts, (4.3) may be expressed in terms of $Si(x)$ and $Ci(x)$ functions. In this way it may be shown that the approximation (4.3) for G and H has the value

$$2\pi A^2U^3K/\alpha^2 = 2\pi\, A^2U\, K/a^2$$

where

$$\begin{aligned} K = {} & (-U^{-1} + \beta^{-1})(1 - \cos U)(1 - \cos \beta) \\ & + (Si\, U - Si\, \beta)(1 + \cos \alpha - 2\alpha^{-1} \sin \alpha) \\ & + (Si\, 2U - Si\, 2\beta)(-\cos \alpha + \alpha^{-1} \sin \alpha) \\ & + (Ci\, 2U - Ci\, 2\,|\,\beta\,| - Ci\, U + Ci\,|\,\beta\,|)(\sin \alpha + \alpha^{-1} \cos \alpha) \\ & + \alpha^{-1}(2 + \cos \alpha)[\log_e (U/|\,\beta\,|) - Ci\, U + Ci\,|\,\beta\,|\,] \end{aligned} \tag{4.5}$$

Here $\alpha = aU$ and $\beta = \alpha - U = (a - 1)U$. When $f = 2f_b$, a has the value 2 and K is zero, as it should be. When $f = f_b$, i.e., when $a = 1$ and $\alpha = U$,

$$\begin{aligned} K = {} & (1 + \cos U - 2U^{-1} \sin U)\, Si\, U \\ & + (-\cos U + U^{-1} \sin U)\, Si\, 2U \\ & + (Ci\, 2U - Ci\, U - \log_e 2)(\sin U + U^{-1} \cos U) \\ & + U^{-1}\,(2 + \cos U)(\log_e U + .577.. - Ci\, U) \end{aligned} \tag{4.6}$$

Values of K are tabulated in Table 4.2 for various frequencies and delays ($a = f/f_b$ and $k = 4U/\pi = 8f_bT$). Cases 3 and 4 in Table 4.1 show that when U is very small $K \approx a^2U^3(2 - a)/8$ and when U is very large $K \approx \pi$, $\pi/2$, or 0 according to whether $a < 1$, $a = 1$, or $a > 1$.

Case 3 in Table 4.1 is a special case of Case 2 which may be obtained

TABLE 4.2 — VALUES OF K

k	$a = 0$	0.25	0.50	0.75	1.00	1.25
0	0	0.0	0.0	0.0	0.0	0.0
1	0	0.006	0.022	0.041	0.058	0.068
2	0	0.048	0.164	0.305	0.422	0.473
3	0	0.141	0.490	0.890	1.20	1.26
4	0	0.286	0.961	1.74	2.19	1.45
5	0	0.453	1.57	2.64	3.05	2.62
6	0	0.673	2.16	3.36	3.41	2.45
7	0	0.897	2.69	3.68	3.13	1.73
8	0	1.14	3.13	3.65	2.40	0.883
9	0	1.39	3.45	3.34	1.62	0.370
10	0	1.66	3.68	2.95	1.20	0.378
11	0	1.93	3.79	2.64	1.32	0.741
12	0	2.20	3.79	2.51	1.89	1.10

by letting U become very small in (4.3). The value of P_I/P_S corresponding to Case 3 has been given by Albersheim and Schafer.[3] Case 4 may be obtained by letting U become large in (4.5) and (4.6).

When U is very small, expression (2.8) for $R_v(\tau)$ becomes

$$R_v(\tau) \approx AU^2u^{-1} \sin u \tag{4.7}$$

Assuming $AU^2 \ll 1$ and substituting (4.7) in the "second order modulation" approximation (2.9) for G and H gives us another derivation of Case 3 (see (A3–2)). However, if the rms frequency deviation of the signal is so large that AU^2 is not small, even though $U \ll 1$, we have Case 5, the case investigated by Lewin.[2] The formulas given in Table 4.1 are obtained when (4.7) is set in the integrals (2.4) and (2.5) for G and H, and the results compared with the definition (A3–1) of $I(b, a)$.

When U is very large, expression (2.8) for $R_v(\tau)$ becomes

$$R_v(\tau) \approx \begin{cases} A[\pi U - 2\cos u - 2uSi(u)], & 0 \leq u < U \\ 0, & U < u \end{cases} \tag{4.8}$$

Substituting (4.8) in (2.9) and integrating by parts twice leads to another derivation of Case 4. The expression for G given in Case 6 is obtained from (4.8) and (2.4) by the method outlined in Appendix II. $A\pi U$ is assumed to be so large that most of the contribution to the value of the integral (2.4) for G comes from the region around $u = 0$. It is also assumed that $R_v(\tau) + 1$ is negligible in comparison with $\exp R_v(\tau)$ in this region. This leads to the approximation

$$G \approx 2\int_0^\infty e^{2A[1-\cos u - uSi(u)]} \cos au \, du \tag{4.9}$$

which holds when $U \gg 1$ and $A\pi U \gg 1$.

The expression for G in Case 6 is merely the leading term in the asymptotic expansion arising from the saddle point at $u = iy_1$. When further terms in this expansion are obtained (using, for example, equation (10.4) of Reference 7) it is found that the expression for Case 6 should be multiplied by

$$1 + \frac{[4cs - 5y_1 + s^2(y_1^{-1} - 2y_1)]}{48As^3} + \cdots. \tag{4.10}$$

where c and s denote cosh y_1 and sinh y_1, respectively. The next term consists of $1/A^2$ times a function of y_1, and so on. When y_1 becomes small, as it does when A becomes large or a becomes small, (4.10) becomes

$$1 + \frac{1}{48A} + \frac{103}{13824A^2} + \cdots.$$

However, comparison of Tables 4.3 and 4.4 shows that the formula of Case 6 gives fairly reliable values of G when A is as small as 0.5 (U must be large, of course).

When $(1 - a)$ is small and $A \ll 1$, but U still large enough to make $A\pi U \gg 1$, (4.9) gives

$$G \approx Aa^{-2}\left[\pi + 2 \text{ arc tan } \frac{(1 - a)}{A\pi}\right] + 0(A^2/a^2) \tag{4.11}$$

This may be obtained by letting A become small in

$$G \approx 4(A/a)^2 \int_0^\infty e^{2Ay} Si(u) F \, du \tag{4.12}$$

$$y = 1 - \cos u - uSi(u)$$

$$F = -2 \cos au Si(u) + Si[(1 - a)u] + Si[(1 + a)u]$$

which may be obtained from (4.9) by integrating by parts twice.

The formula for G given in Case 7, Table 4.1, is obtained when U is taken to be of order unity and A is assumed to be so large that only the exponential term in the integrand of (2.4) is of importance. Most of the contribution comes from around $u = 0$ where

$$R_v(\tau) = R_v(0) - u^2A(1 - U^{-1} \sin U) + \cdots.$$

When $U \gg 1$, and $A \gg 1$, Cases 6 and 7 both give

$$G \approx (\pi/A)^{1/2} \exp [-a^2/(4A)]$$

which leads to an expression for P_I/P_S similar to one given by Alber-

TABLE 4.3 — VALUES OF G AND H OBTAINED BY NUMERICAL INTEGRATION

	$a = 0$	0.25	0.50	0.75	1.00	1.25
G for $U = 3$						
$A = 0.125$	0.690	0.651	0.599	0.466	0.326	0.224
0.25	1.53	1.46	1.29	1.06	0.808	0.561
0.50	2.17	2.09	1.97	1.63	1.38	0.993
H for $U = 3$						
$A = 0.125$	0.424	0.393	0.335	0.260	0.182	0.111
0.25	0.574	0.528	0.447	0.339	0.233	0.136
0.50	0.283	0.255	0.211	0.156	0.104	0.039
G for $U = 6$						
$A = 0.125$	2.62	2.42	1.85	1.18	0.631	0.257
0.25	3.26	3.05	2.50	1.81	1.17	0.695
0.50	2.55	2.46	2.21	1.86	1.47	1.10
H for $U = 6$						
$A = 0.125$	0.802	0.705	0.465	0.227	0.067	0.0031
0.25	0.266	0.231	0.148	0.064	0.0126	0.0035
0.50	0.0092	0.0079	0.0048	0.0018	0.00017	0.00019
	$A = 0.0625$	0.125	0.25	0.50		
G for $a = 1$						
$U = 1.5$	0.0126	0.045	0.155	0.446		
3	0.111	0.326	0.808	1.38		
6	0.245	0.631	1.17	1.47		
12	0.300	0.657	1.16	1.47		

sheim and Schafer[3] for long delay. When $U \ll 1$ and $AU^2 \gg 1$, Cases 5 and 7 both give

$$G \approx (6\pi A^{-1} U^{-2})^{1/2} \exp [-3a^2 A^{-1} U^{-2}/2]$$

Before the approximations listed in Table 4.1 were developed, a number of values of G and H were obtained from (2.4) and (2.5) by numerical integration. These values, given in Table 4.3, are the best we have and may be used to check the various approximations.

As an example of the values given by our approximations we take the case $U = 6$. In Table 4.4, "G — Case 6" has been computed from the formula given in Case 6, Table 4.1 (which assumes $U \to \infty$). When these values are compared with the corresponding ones in Table 4.3, it is seen that the agreement is not good for $A = 0.125$. Better agreement is shown by "G — Improved Case 6" in which the values are computed from (2.4) and (4.8) by the method of Appendix II. It differs from "Case 6" in that $F(u)$ of (A2–1) is $\exp [R_v(\tau)] - R_v(\tau) - 1$ instead of merely $\exp [R_v(\tau)]$.

TABLE 4.4 — APPROXIMATE VALUES OF G FOR $U = 6$

	$a = 0$	0.25	0.50	0.75	1.00	1.25
G—Case 6						
$A = 0.125$	5.01	4.13	2.54	1.32	0.63	0.28
0.25	3.54	3.26	2.58	1.80	1.14	0.66
0.50	2.51	2.42	2.17	1.82	1.43	1.06
G—Improved Case 6						
$A = 0.125$	2.62	2.37	1.74	1.06	0.56	0.26
0.25	3.17	2.94	2.39	1.72	1.10	0.66
0.50	2.50	2.41	2.17	1.82	1.42	1.06

5. INTERCHANNEL INTERFERENCE POWER

The values of P_I/P_S, the ratio of the interchannel interference power to the signal power, may be computed from the formulas (2.10) and (2.11) when G and H are known. One would like to have curves giving P_I/P_S for representative combinations of echo delay, signal power, and channel position which are likely to occur in practice. However, the large number of such combinations coupled with the difficulty of computing G and H leads us to restrict ourselves mostly to curves for Cases 2 and 6 in Tables 3.1 and 4.1. In all cases the signal power P_S (per cps) is taken to be equal to the constant value P_0 (measured in (radians)^2/cps for PM and in (radians/sec)^2/cps for FM) over the signal band $(0, f_b)$, and is zero outside this band.

Case 2 is the "second order modulation" approximation which, roughly speaking, applies when the echo delay is very short or when the rms deviation of the phase angle (for PM) or of the frequency (for FM) is small. Case 6 applies when the echo delay is very long.

(*a*) *"Second order modulation" approximation for PM — Table 3.1, Case 2.* Since $G \approx H$, equation (2.10) may be written as

$$\begin{aligned} 10 \log_{10} (P_I/P_S) &\approx \rho + D_1 + D_2 + D_3 \\ \rho &= 10 \log_{10} r^2 \\ D_1 &= 10 \log_{10} (1 - \cos 2pT) \\ D_2 &= 10 \log_{10} (P_0 f_b) \\ D_3 &= 10 \log_{10} J \end{aligned} \tag{5.1}$$

where ρ is the reflection coefficient expressed in decibels, D_1 is a quantity which varies rapidly with T and whose representative value is zero. J is the function of a and U defined by equation (3.1). The quantities P_0,

f_b, a, U are defined by equation (2.1) and Case 1 of Table 3.1. $P_0 f_b$ is the average signal power in (radians)^2, $a = f/f_b$ gives the channel position and $U = 2\pi f_b T$ measures the echo delay.

The approximation (5.1) holds when $2P_0 f_b(1 - U^{-1} \sin U)$ is small in comparison with unity.

Fig. 5.1 shows D_3 plotted as a function of $f_b T$ for various values of a. The values of J which were used were taken from Table 3.2. Values of J for $U \ll 1$ may be obtained from the expression given for G in Case 3, Table 3.1. Case 4 gives another special case.

(*b*) *Large delay, $U \gg 1$ for PM — Table 3.1, Case 6.* When U is very large, G is a rapidly oscillating function of a. When in addition $P_0 f_b$ is small, Case 4 shows that J fluctuates between $(1 - a/2)/2$ and $3(1 - a/2)/2$. The corresponding fluctuations in D_3 are noticeable in Fig. 5.1 for the larger values of $f_b T$. If $2P_0 f_b$ is large compared to unity (and the delay is large), equation (5.1) no longer holds. In this case $G \gg H$ and we may write (2.10) as

$$\begin{aligned} 10 \log_{10} (P_I/P_S) &\approx \rho + D_4 \\ \rho &= 10 \log_{10} r^2 \\ D_4 &= 10 \log_{10} (2\pi P_0 f_b)^{-1} G \\ G &\approx e^{-b_0}[I(b_0, a) + 2I(b_1, a) \cos aU] \end{aligned} \tag{5.2}$$

where b_0 and b_1 are defined in Case 6, Table 3.1. It is seen that as a increases from 0 to 1, D_4 oscillates rapidly between limits D_4^+ and D_4^- corresponding to G^+ and G^- which are defined by equation (3.4) and tabulated in Table 3.3. Table 5.1 gives values of D_4^+ and D_4^- computed from Table 3.3.

The entries corresponding to the values 0.25 and 0.50 for $P_0 f_b$ must be used with caution in equation (5.2) since they do not satisfy $2P_0 f_b \gg 1$ and H is not negligible in comparison with G.

(*c*) *"Second order modulation" approximation for FM — Table 4.1, Case 2.* By making use of the expressions for G and H given in Case 2 we may write equation (2.11) as

$$\begin{aligned} 10 \log_{10}(P_I/P_S) &\approx \rho + D_1 + D_2' + D_3' \\ D_2' &= 10 \log_{10} A \\ D_3' &= 10 \log_{10} UK \end{aligned} \tag{5.3}$$

where ρ and D_1 are defined by (5.1) and A by (2.8). K is the function of a and U defined by (4.5). A is proportional to the signal power: $A =$

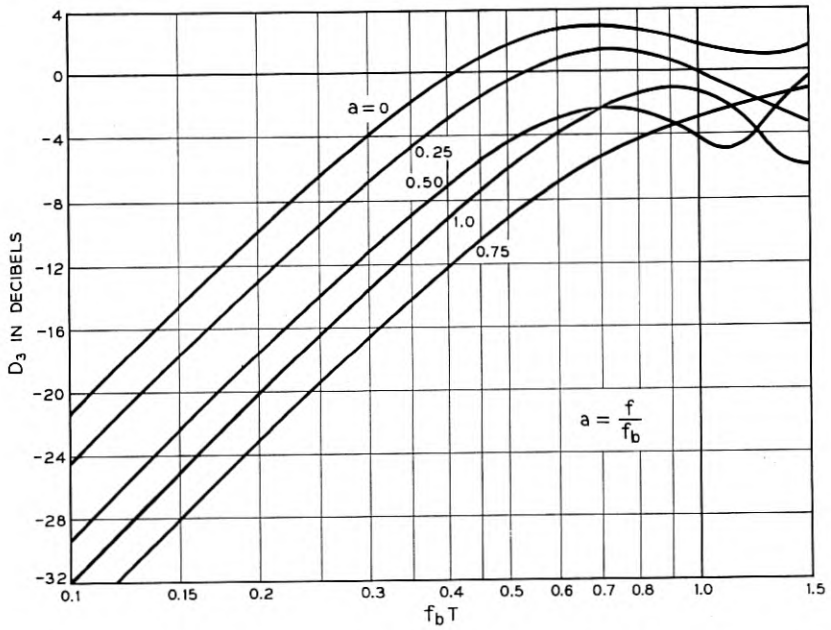


Fig. 5.1 — For "Second Order Modulation" the ratio P_1/P_S for PM depends upon D_3 as shown by equation (5.1).

TABLE 5.1 — VALUES OF D_4^+ AND D_4^-. THE UPPER NUMBER OF AN ENTRY IS D_4^+ AND THE LOWER ONE IS D_4^-

	$a = 0$	0.25	0.50	0.75	1.00—
$P_0 f_b$ (radians)^2					
0.25	−6.12	−6.68	−7.31	−8.07	−9.01
	−9.92	−10.37	−10.96	−11.67	−12.56
0.50	−4.89	−5.40	−6.00	−6.75	−7.67
	−7.94	−8.31	−8.79	−9.45	−10.34
0.75	−4.83	−5.31	−5.87	−6.57	−7.47
	−7.24	−7.56	−8.01	−8.63	−9.46
1.00	−5.20	−5.61	−6.13	−6.78	−7.67
	−7.12	−7.38	−7.77	−8.33	−9.17
2.00	−7.65	−7.89	−8.25	−8.74	−9.44
	−8.30	−8.44	−8.74	−9.17	−9.82
4.00	−11.99	−12.08	−12.25	−12.55	−12.94
	−12.05	−12.13	−12.31	−12.55	−12.94

$(\sigma/f_b)^2$ where σ is the rms frequency deviation of the signal in cycles per second. D_2' and D_2 play similar roles in (5.3) and (5.1).

The approximation (5.3) holds when $2A[USi(U) - 1 + \cos U]$ is small in comparison with unity.

The values of K given in Table 4.2 lead to the curves for D_3' shown in Fig. 5.2. When $U \ll 1$, Case 3 shows that $D_3' \approx 10 \log_{10} [a^2U^4(2 - a)/8]$, and Case 4 shows that when $U \gg 1$ (provided $a < 1$ and $A\pi U \ll 1$) $D_3' \approx 10 \log_{10} (\pi U) = 12.95 + 10 \log_{10} f_bT$.

(*d*) *Large delay, $U \gg 1$ for FM — Table 4.1, Case 6.* It has just been pointed out that when U becomes very large, P_I/P_S depends upon the delay only through the term $D_3' \approx 10 \log_{10} \pi U$ (neglecting the rapidly varying term D_1) if $AU\pi \ll 1$. If $AU\pi \gg 1$, P_I/P_S becomes independent of U as $U \to \infty$. This follows from the fact that the formulas of Case 6 allow us to write (2.11) as

$$\begin{aligned} 10 \log_{10} (P_I/P_S) &= \rho + D_4' \\ D_4' &\approx 10 \log_{10} [Ga^2(2\pi A)^{-1}] \end{aligned} \tag{5.4}$$

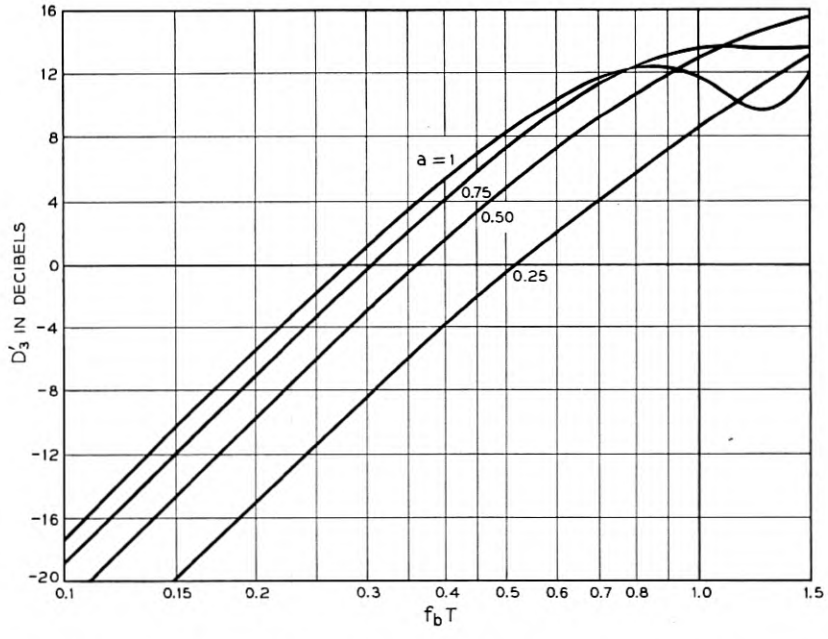


Fig. 5.2 — For "Second Order Modulation" the ratio P_I/P_S for FM depends on D_3' as shown by equation (5.3).

where ρ is defined by (5.1) and G depends only on a and A through

$$G \approx [\pi y_1/A \sinh y_1]^{1/2} \exp [-2A (\cosh y_1 - 1)]$$
$$\frac{a}{2A} = \int_0^{y_1} \frac{\sinh v}{v}\, dv \tag{5.5}$$

Fig. 5.3 shows D_4' plotted as a function of A for various values of a. It is assumed that $A\pi U \gg 1$.

(e) *Small delay and large rms frequency deviation for FM — Table 4.1, Cases 5 and 7.* It turns out that Case 5 (Lewin's, case, $U \ll 1$) and Case 7 ($A \gg 1$ and U of order unity) may be combined into a single case by taking the quantity b in the formulas of Case 5 to be $6A(1 - U^{-1} \sin U)$ instead of AU^2. When $U \ll 1$, Case 5 is obtained. When $A \gg 1$ the asymptotic expansion for $I(b, a)$ leads to Case 7 if U is 0(1).

In order to put this combined case in a form suited to calculation we write (2.11) as

$$\frac{P_I}{P_s} = \frac{r^2a^2G}{2\pi A}\left(1 - \frac{H}{G}\cos 2pT\right)$$
$$\approx \frac{r^2 6(1 - U^{-1}\sin U)a^2e^{-b}I(b, a)}{2\pi b}\left(1 - \frac{H}{G}\cos 2pT\right) \tag{5.6}$$
$$10 \log_{10} (P_I/P_s) \approx \rho + D_1' + D_5' + D_6'$$

where ρ is given by (5.1) and

$$\begin{aligned} D_1' &= 10 \log_{10} (1 - (H/G) \cos 2pT) \\ D_5' &= 10 \log_{10} (1 - U^{-1} \sin U) \\ D_6' &= 10 \log_{10} 6a^2\, e^{-b}\, I(b, a)/(2\pi b) \\ b &= 6A(1 - U^{-1} \sin U) \end{aligned} \tag{5.7}$$

Fig. 5.4 shows values of D_6', computed from the values of $I(b, a)$ given in Appendix III, plotted as a function of b for various values of a. The maximum value of 3 db for D_1' occurs when $AU^2 \ll 1$ and $\cos 2pT = -1$. When AU^2 is large D_1' is approximately zero.

Fig. 5.5 and 5.6 show, in a rough way, the regions in which the various approximations apply. For PM, the delay and the rms phase deviation (measured by $U = 2\pi f_b T$ and $(P_0 f_b)^{1/2}$, respectively) are the parameters which determine the type of approximation to be used. The regions in the $[(P_0 f_b),^{1/2} U]$ plane shown in Fig. 5.5 are marked with the numbers 2a, 3, 4, 5, 6b where the integer indicates the case number in Table 3.1 and the letters a and b refer to Cases (a) and (b) in this section. Fig. 5.6 is

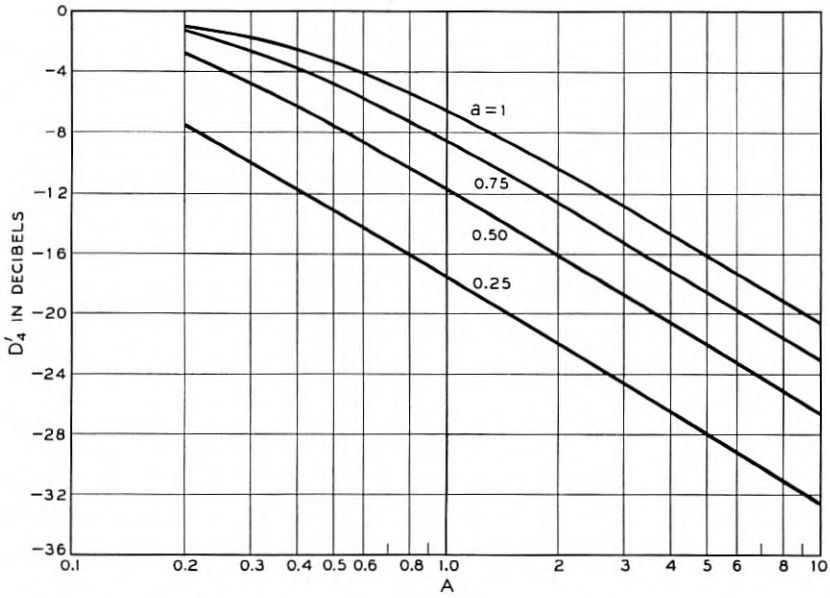


Fig. 5.3 — For long delayed echoes the ratio P_I/P_S for FM depends upon D_4' as shown by equation (5.4).

the corresponding figure for FM. The coordinates are U and $A^{1/2}$, where $A^{1/2}(=\sigma/f_b)$ measures the rms frequency deviation. The region numbers are 2c, 3, 4, 5, 6d, 7e where the integers indicate the case number in Table 4.1. It will be noted that there are regions where no approximation is available. However, an answer may always be obtained by numerical integration of equations (2.4) and (2.5) for G and H.

Fig. 5.7 shows values of P_I/P_S for the top channel ($a = 1$) where the interference is often at a maximum in an FM system. The coordinates $(A^{1/2}, f_bT)$ are essentially the same as those of Fig. 5.6. In order to simplify the plotting, the phase angle $2pT$ is assumed to be such that $\cos 2pT$ is zero so that the contours are given by

$$\text{Constant} = 10 \log_{10}\left(\frac{P_I}{r^2 P_S}\right) = 10 \log_{10}\left(\frac{G}{2\pi A}\right)$$

The contours have been obtained in part from the various approximations where applicable and in part from values obtained by numerical computation from the exact expression. While there are, of necessity, some areas of uncertainty in Fig. 5.7, it should be adequate for most engineering purposes. No corresponding curves have been computed for the case of phase modulation.

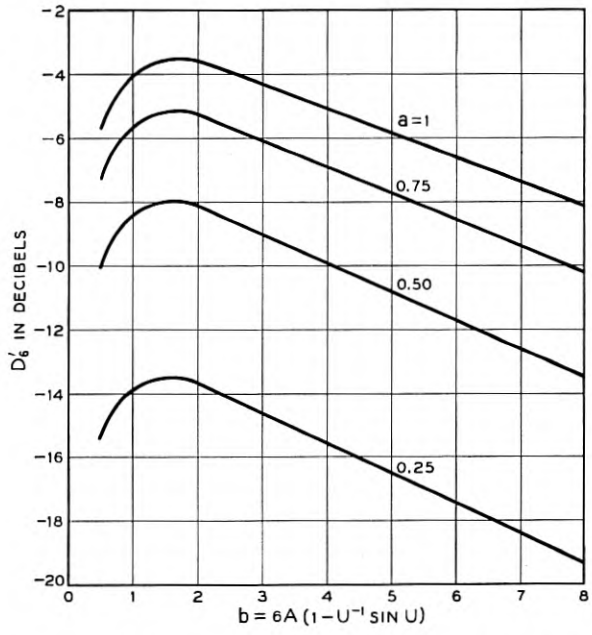


Fig. 5.4 — Under the condition of large rms frequency deviation and small delay the ratio P_I/P_S for FM depends upon D_6' as shown by equation (5.6).

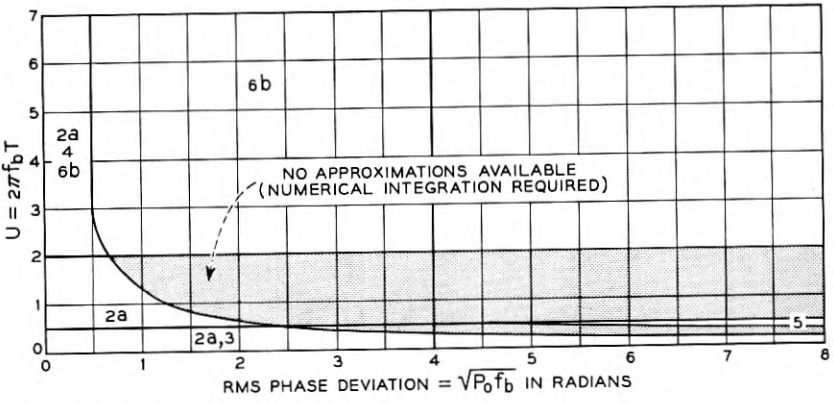


Fig. 5.5 — Regions of validity for the various approximations for PM. The integers refer to case numbers in Table 3.1 and the letters to cases discussed in Section 5. This figure and Fig. 5.6 are intended to give only an idea of the relative positions of the regions.

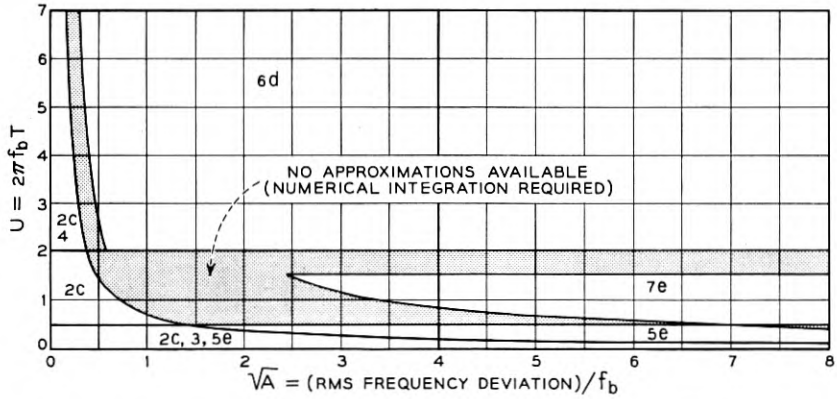


Fig. 5.6 — Regions of validity for the various approximations for FM. The integers refer to case numbers in Table 4.1 and the letters to cases discussed in Section 5.

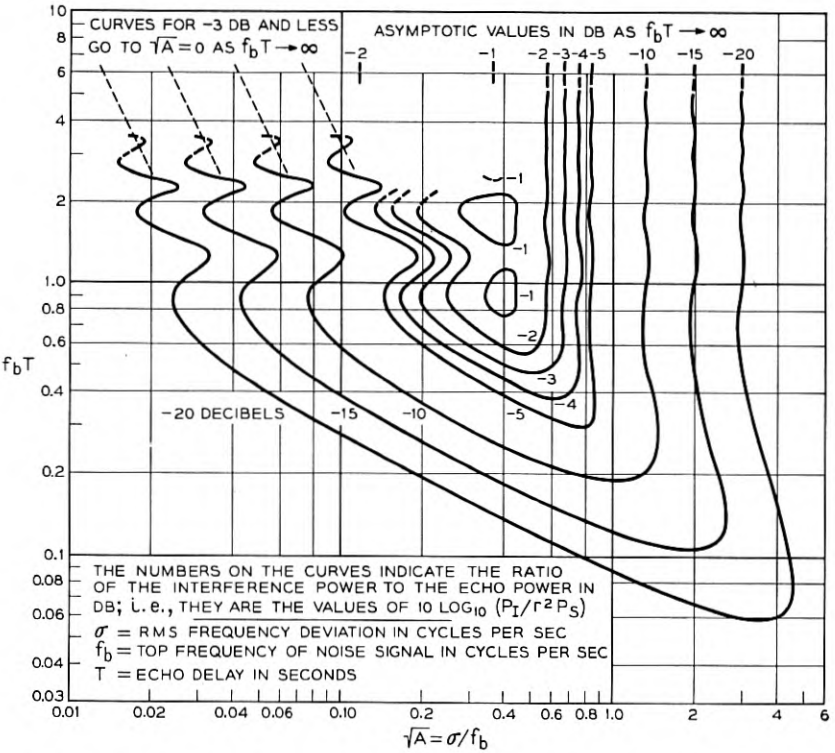


Fig. 5.7 — Contours of constant interference in the top channel of a multichannel FM system.

It should be noticed that Fig. 5.7 is plotted for the case $a = f/f_b = 1$. If Fig. 5.7 were plotted for values of a slightly less than unity there would not be much change except in the upper left hand corner (A small and f_bT large), where the interference would tend to be 3 db stronger. This discontinuous behavior as a passes through unity is shown by Case 4, Table 4.1 where $U \gg 1$ and $A\pi U \ll 1$. When $U \gg 1$ and $A\pi U \gg 1$, as occurs when $f_bT \to \infty$ (with A held fixed), equation (4.11) gives

$$10 \log_{10} \frac{P_I}{r^2 P_S} \approx 10 \log_{10} \left[\frac{1}{2} + \frac{1}{\pi} \arctan \left(\frac{1 - a}{A\pi} \right) \right]$$

when $A \ll 1$. This shows that at $f_bT = \infty$, the discontinuity arises for values of A near $(1 - a)/\pi$. When $A \gg 1$ and $U \gg 1$, we have

$$10 \log_{10} \frac{P_I}{r^2 P_S} \approx -15 \log_{10} A - 10 \log_{10} \frac{2\pi^{1/2}}{a^2} - \frac{4.34a^2}{4A}$$

which changes only slowly as $a \to 1$.

6. USE OF EQUIVALENT ECHO TO ESTIMATE INTERCHANNEL INTERFERENCE

When a steady sinusoid $\exp(i\omega t)$ is applied to a transmission medium of the sort we have under consideration, the output is $\exp(i\omega t - \alpha - i\beta)$ where α and β are the attenuation and phase shift, respectively. Distortionless transmission occurs when α is constant and β has a constant slope over the essential range of frequencies. Departures from these ideal conditions cause interchannel interference in multichannel FM and PM systems. Our evaluation of the interference caused by a small echo may equally well be regarded as an evaluation of interference for a particular kind of amplitude and phase distortion, namely that given by

$$\alpha \approx -r \cos \omega T, \qquad \beta \approx r \sin \omega T. \tag{6.1}$$

These expressions are obtained by writing $\exp(i\omega t) + r \exp[i\omega(t - T)]$ in the form $\exp(i\omega t - \alpha - i\beta)$ when $|r| \ll 1$.

The analysis given by E. D. Sunde in Section 1 of Reference 8 shows that a minimum phase system in which $\beta = r \sin \omega T$ also has $\alpha = -r \cos \omega T$ as in (6.1). This suggests a procedure for calculating interchannel interference from phase data alone when (1) the distortion is known to be of the minimum phase type and (2) the variation of phase with frequency can be approximated by a sine function. In such cases we can apply our echo analysis directly by identifying r as the amplitude of the phase oscillation and T as the reciprocal of the period.

In carrier multiplex systems the sinusoidal approximation need hold only in the region around the carrier frequency f_0 where $2\pi f_0 = p$ and p is the radian frequency appearing in equation (1.1). For FM we shall, in this section, arbitrarily take the region to extend from $f_0 - 4\sigma$ to $f_0 + 4\sigma$ where σ is the rms frequency deviation of the signal. For PM and the signal power spectrum given by equation (2.1) we may take the region to be $f_0 \pm 4f_b(P_0 f_b/3)^{1/2}$.

A special case occurs when the nonlinear portion of β may be represented as $a_2(f - f_0)^2/2$ in the region of interest. We can think of $r \sin \omega T$ as going through several oscillations between $f = 0$ and $f = f_0$, and that a maximum, if $a_2 < 0$, (or a minimum, if $a_2 > 0$) of $r \sin \omega T$ occurs at $f = f_0$, i.e. at $\omega = p$. The band of interest is taken to be narrow enough to lie in the immediate vicinity of the maximum. This sort of curve fitting is permissible since equation (2.3) shows that the interchannel interference depends on the carrier frequency only through the term $\cos 2pT$. Furthermore, constant terms and terms linear in frequency in the expression for β do not affect the amount of interchannel interference.

At the maximum mentioned in the preceding paragraph $\omega = p$, $\sin pT = 1$ and $\cos 2pT = -1$. Near this maximum $r \sin \omega T$ is

$$r \cos 2\pi(f - f_0)T \approx r - r4\pi^2(f - f_0)^2T^2/2 \tag{6.2}$$

In order that this approximation may hold over the region $f_0 \pm 4\sigma$ (for FM), we require

$$2\pi(4\sigma)T \leqq 1$$

We take T to be as large as possible, namely

$$T = 1/8\pi\sigma \tag{6.3}$$

in order to make r as small as possible, since our work assumes $r \ll 1$. Comparison of $a_2(f - f_0)^2/2$ with (6.2) gives

$$r = -a_2(2\pi T)^{-2} = -16a_2\sigma^2 \tag{6.4}$$

which must be small compared to unity if our results are to be used. This result holds for $a_2 < 0$. When $a_2 > 0$ expression (6.3) still holds for T but now $r = 16a_2\sigma^2$. Similar expressions hold for PM. These values of r and T may now be inserted in our formulas to determine the interchannel interference.

From the definitions of A and U it may be shown that (6.3) is equivalent to $AU^2 = 1/16$. Therefore the "second order modulation" approximation given by Case 2 in Table 4.1 may be used. Also $G - H \cos 2pT$ is

approximately equal to $2G$ because $\cos 2pT = -1$. It turns out that the second order modulation approximation may also be used in the PM case.

When a_2 is sufficiently small, considerations such as those above show that the ratio of the interference power for $\beta = a_2(f - f_0)^2/2$ (radians) to the received signal power at the frequency f is

$$\begin{aligned} P_I/P_S &= (a_2\sigma f_b/2)^2\, a^2(2 - a) \text{ for FM} \\ P_I/P_S &= (a_2 f_b^2/2)^2(P_0 f_b)(12 - 30a + 20a^2 - a^5)/30 \text{ for PM} \end{aligned} \tag{6.5}$$

Here $a = f/f_b$ where $(0, f_b)$ is the frequency band of the signal. The rms frequency deviation of the signal for FM is σ cps, and the rms phase deviation of the signal for PM is $(P_0 f_b)^{1/2}$ radians. The first equation in (6.5) comes from a special case of Case 2, namely, Case 3 of Table 4.1, and requires the additional assumption $f_b/4\sigma \ll 1$ (corresponding to $U \ll 1$). The second equation requires a similar additional assumption.

Appendix I

Derivation of a General Theorem on the Interchannel Interference Spectrum

Let $w_a(f)$ be a finite power spectrum having limited total fluctuation for $0 \leqq f \leqq \infty$. Let the total power be finite so that the integral of $w_a(f)$ from $f = 0$ to $f = \infty$ converges absolutely. We define two auxiliary spectra

$$\begin{aligned} w_\epsilon(f) &= \begin{cases} w_a(f), & |f - f_0| < \epsilon \\ 0, & |f - f_0| > \epsilon \end{cases} \\ w_b &= w_a(f) - w_\epsilon(f) \end{aligned} \tag{A1-1}$$

and note that the autocorrelations corresponding to these spectra must satisfy

$$R_b(\tau) = R_a(\tau) - R_\epsilon(\tau) \tag{A1-2}$$

We consider the problem of transmitting the ensemble having the spectrum $w_b(f)$ through a system in which the input and output autocorrelation functions, $\Psi_1(\tau)$ and $\Psi_2(\tau)$, respectively, are related by

$$\Psi_2(\tau) = F[\Psi_1(\tau)]$$

Here $F(z)$ and its derivatives $F'(z)$ and $F''(z)$ are assumed to be finite continuous functions which exist over the range of z of interest.

Let the output spectrum corresponding to the input spectrum $w_b(f)$ be $w_B(f)$. We shall show that as $\epsilon \to 0$ the value of $w_B(f)$ in the range

$|f - f_0| < \epsilon$ approaches

$$4\int_0^\infty \{F[R_a(\tau)] - F'(0)R_a(\tau)\} \cos 2\pi f\tau \, d\tau \tag{A1-3}$$

When we multiply this expression by 2ϵ and set $f = f_0$, we obtain the power appearing at the output of the system in an unloaded channel of width 2ϵ centered on f_0. Here we are not interested in values of $w_B(f)$ outside the range $|f - f_0| < \epsilon$.

First we note several properties of autocorrelation functions. From

$$R(\tau) = \int_0^\infty w(f) \cos 2\pi f\tau \, df$$

it follows that $|R(\tau)| \leqq R(0)$. Also, if $w(f)$ has limited total fluctuation in the interval $(0, \infty)$, the Riemann-Lesbesgue lemmas* and the absolute convergence of the integral for $R(0)$ show that $R(\tau) = 0(1/\tau)$ as $\tau \to \infty$. Thus we may find positive numbers A, B, C such that for $0 < \tau$ and any ϵ less than some fixed value

$$\begin{aligned} &|R_a(\tau)| < A/\tau \\ &|R_\epsilon(\tau)| \leqq R_\epsilon(0) = \int_{f_0-\epsilon}^{f_0+\epsilon} w_a(f)\, df < B\epsilon \\ &|R_\epsilon(\tau)| < C/\tau \end{aligned} \tag{A1-4}$$

By the extended theorem of the mean the autocorrelation function corresponding to $w_B(f)$ is

$$\begin{aligned} R_B(\tau) &= F[R_b(\tau)] = F[R_a(\tau) - R_\epsilon(\tau)] \\ &= F[R_a(\tau)] - F'[R_a(\tau)]R_\epsilon(\tau) + r \\ |r| &= 2^{-1}R_\epsilon^2(\tau)\,|F''[R_a(\tau) - \theta R_\epsilon(\tau)]| < R_\epsilon^2(\tau)D \end{aligned} \tag{A1-5}$$

where $0 \leqq \theta \leqq 1$ and D is a positive number such that $|F''(z)| < D$. Then

$$\begin{aligned} w_B(f) &= 4\int_0^\infty R_B(\tau) \cos 2\pi f\tau \, d\tau = I_1 - I_2 + I_3 \\ I_1 &= 4\int_0^\infty F[R_a(\tau)] \cos 2\pi f\tau \, d\tau \\ I_2 &= 4\int_0^\infty F'[R_a(\tau)]R_\epsilon(\tau) \cos 2\pi f\tau \, d\tau \\ I_3 &= 4\int_0^\infty r \cos 2\pi f\tau \, d\tau \end{aligned} \tag{A1-6}$$

* See, for example, Whittaker and Watson, Modern Analysis, 4th edition, p. 172.

Since $F'[R_a(\tau)] = F'(0) + s$, where by the mean value theorem

$$| s | = | R_a(\tau)\, F''[\theta R_a(\tau)] | < | R_a(\tau) | D$$

$$0 \leqq \theta \leqq 1$$

I_2 may be written as the sum of two integrals, the second of which has an absolute value not greater than

$$4 \int_0^\infty | R_a(\tau) D R_\epsilon(\tau) | \, d\tau = 4D \int_0^T | R_a(\tau) R_\epsilon(\tau) | \, d\tau$$

$$+ 4D \int_T^\infty | R_a(\tau) R_\epsilon(\tau) | \, d\tau < 4D R_a(0) B \epsilon T$$

$$+ 4D \int_T^\infty A C \tau^{-2} \, d\tau = 4D[R_a(0) B \epsilon T + AC/T]$$

where T is an arbitrary number and we have used the inequalities (A1-4). Choosing $T = \epsilon^{-1/2}$ shows that the last expression is $0(\epsilon^{1/2})$. Hence

$$I_2 = 4F'(0) \int_0^\infty R_\epsilon(\tau) \cos 2\pi f \tau \, d\tau + 0(\epsilon^{1/2})$$

$$= 0(\epsilon^{1/2}) + F'(0) \begin{cases} w_a(f), & | f - f_0 | < \epsilon \\ 0, & | f - f_0 | > \epsilon \end{cases}$$

Therefore in the range $| f - f_0 | < \epsilon$, which comprises the only frequencies of interest in the channel interference spectrum,

$$I_2 = 4F'(0) \int_0^\infty R_a(\tau) \cos 2\pi f \tau \, d\tau + 0(\epsilon^{1/2})$$

By use of the inequalities for $| r |$ and $| R_\epsilon(\tau) |$ we see that

$$| I_3 | < 4 \int_0^\infty R_\epsilon^2(\tau) D d\tau < 4D \int_0^T B^2 \epsilon^2 d\tau + 4D \int_T^\infty C^2 \tau^{-2} d\tau$$

$$= 4D(B^2 \epsilon^2 T + C^2/T)$$

If we choose $T = 1/\epsilon$ this expression is $0(\epsilon)$.

When we collect our results and let ϵ become vanishingly small, we see that expression (A1-6) for $w_B(f)$ approaches (A1-3) for frequencies in the range $| f - f_0 | < \epsilon$. Thus in the limit as $\epsilon \to 0$ we may use the autocorrelation function

$$F[R_a(\tau)] - F'(0) R_a(\tau)$$

to compute the interchannel interference spectrum. This is the result used in (1.22).

Appendix II

APPROXIMATE EVALUATION OF INTEGRALS OF A CERTAIN TYPE

The problem of evaluating the integral G defined by (2.4) is quite a difficult one. Here we shall outline a method which often may be used to obtain an idea of the order of the magnitude of such an integral.

Let $F(u)$ be an even analytic function of u such that the major contribution to the value of

$$I(a) = 2\int_0^\infty F(u)\cos au\,du = \int_{-\infty}^{\infty} F(u)e^{iau}\,du \qquad \text{(A2-1)}$$

comes from a saddle point on the positive imaginary u axis. Then the "method of steepest descents" suggests that an approximate value of $I(a)$ may be obtained by the following procedure.

1. Set $f(y) = F(iy)$ and plot $z = d[\log f(y)]/dy = f'(y)/f(y)$ as a function of y.

2. Draw the horizontal line $z = a$. Suppose its first intersection with the curve obtained in step 1 is at $y = y_1$, and let the slope of the curve, determined either graphically or by differentiation, be $(dz/dy)_{y_1}$ at y_1.

3. Then $I(a) \approx [2\pi/(dz/dy)_{y_1}]^{1/2} f(y_1)e^{-ay_1}$ (A2-2)

It should be noted that (A2-2) cannot be used indiscriminately. Thus, it does not work well for $F(u) = 1/(1 + u^4)$ because there is no saddle point on the imaginary u-axis. However, when it is applied to integrals of the type encountered in our study it appears to do fairly well, as Table 4.4 shows.

Appendix III

LEWIN'S INTEGRAL

Here we study the integral

$$I(b, a) = \int_{-\infty}^{\infty} [e^{bu^{-1}\sin u} - 1 - bu^{-1}\sin u]\cos au\,du \qquad \text{(A3-1)}$$

which occurs in several limiting cases in our work and which has been studied by Lewin[2] for $a = 0$ and $a = 1$.

When the exponential term is expressed as a power series in $(b \sin u)/u$ and the result integrated termwise, we obtain

$$I(b, a) = \sum_{n=2}^{\infty} A_n b^n/n! \tag{A3-2}$$

$$A_n = \int_{-\infty}^{\infty} \left(\frac{\sin u}{u}\right)^n \cos au\, du$$

$$= \frac{2\pi}{2^n(n-1)!} \sum_{m=0} (-)^m C_m{}^n (n - 2m + a)^{n-1}$$

$$A_n \sim (6\pi/n)^{1/2} \exp\,[-3a^2/(2n)]$$

where $C_m{}^n = n!/m!(n\text{-}m)!$ and the last term in the summation for A_n is the last one for which $n - 2m + a$ is positive (assuming $a \neq$ integer) and for which $m \leqq n$. When $n \geqq 2$, A_n is a continuous function of a. Tables III A and III B were computed from (A3-2).

When b is small, the first term in (A3-2) gives, for $0 \leqq a \leqq 2$,

$$I(b, a) \approx b^2\pi(2 - a)/4 \tag{A3-3}$$

and when b is a large positive number the contribution of the exponential term in the region around $u = 0$ gives

$$I(b, a) \approx (6\pi/b)^{1/2} \exp\left(b - \frac{3a^2}{2b}\right) \tag{A3-4}$$

Lewin has given more careful approximations for the $a = 0$ and $a = 1$ cases.

When b is large and negative most of the contribution comes from around $u = \pm 3\pi/2$ where $\exp\,[(b \sin u)/u]$ attains its largest values.

It is found that

$$I(-\beta, a) = 10.76\, \beta^{-1/2} \cos\,(4.49a) e^{0.217\beta - 2.30a^2\beta^{-1}} + R \tag{A3-5}$$

where $\beta = -b$ is a large positive number and R is a remainder term. The numbers in (A3-5) are related to the value $u_0 = 4.493\ldots$ where $(\sin u)/u$ has a minimum.

Computation shows that the value $\beta = 8$ is not large enough to make the leading term in (A3-5) a good approximation for $I(-\beta, a)$. In order to obtain a better approximation we write

$$I(-\beta, a) \approx 2 \int_0^y \exp\,[-\beta u^{-1} \sin u] \cos au\, du$$

$$+ 2 \int_0^y [-1 + \beta u^{-1} \sin u] \cos au\, du \tag{A3-6}$$

$$+ \int_y^{\infty} (\beta u^{-1} \sin u)^2 \cos au\, du$$

TABLE III A — $e^{-b}I(b, a)$ FOR $b > 0$

b	e^b	$a = 0$	0.25	0.50	0.75	1.00	1.25
0	1	0.0	0.0	0.0	0.0	0.0	0.0
0.5	1.649	0.272	0.241	0.209	0.176	0.142	0.107
1.0	2.718	0.761	0.685	0.602	0.511	0.414	0.314
2.0	7.389	1.560	1.440	1.291	1.117	0.919	0.713
3.0	20.08	1.913	1.801	1.645	1.448	1.215	0.968
4.0	54.60	1.974	1.888	1.751	1.566	1.341	1.098
5.0	148.4	1.905	1.844	1.731	1.571	1.372	1.153
6.0	403.4	1.794	1.751	1.660	1.525	1.356	1.166
7.0	1097.	1.680	1.649	1.575	1.463	1.320	1.157
8.0	2981.	1.576	1.552	1.492	1.398	1.277	1.138

TABLE III B — $I(b, a)$ FOR $b < 0$

b	$a = 0$	.25	.50	.75	1.0	1.25
0	0.0	0.0	0.0	0.0	0.0	0.0
−0.5	0.349	0.300	0.254	0.210	0.167	0.125
−1.0	1.25	1.06	0.885	0.723	0.576	0.432
−2.0	4.16	3.41	2.76	2.20	1.76	1.34
−3.0	8.03	6.37	4.97	3.88	3.14	2.46
−4.0	12.6	9.66	7.23	5.49	4.55	3.74
−5.0	17.8	13.2	9.40	6.89	5.93	5.19
−6.0	23.6	16.8	11.4	8.00	7.25	6.85
−7.0	30.0	20.7	13.1	8.71	8.48	8.78
−8.0	37.2	24.8	14.5	8.93	9.59	11.0

where y is such that the quantity within the brackets in (A3-1), with $b = -\beta$, is approximately $(\beta u^{-1} \sin u)^2/2$ when $u > y$. For rough work we may take $y = \beta$. The leading term in (A3-5) arises from the contribution of the region around $u = 3\pi/2$ to the value of the first integral in (A3-6). The contributions from the regions around $u = 7\pi/2$, $11\pi/2$, . . . (if y is large enough) add to the value of R but they are generally small in comparison with the leading term in (A3-5).

Thus we are led to approximate R in (A3-5) by the sum of the second and third integrals (expressed in terms of integral sines and cosines) in (A3-6) with $y = \beta$. When R is replaced by this sum, expression (A3-5) gives values for $b = -8$ which agree fairly well with those in the table.

REFERENCES

1. S. T. Meyers, Nonlinearity in Frequency-Modulation Radio Systems Due to Multipath Propagation, Proc. I.R.E., **34,** pp. 256–265, 1946.
2. L. Lewin, Interference in Multi-Channel Circuits, Wireless Engineer, **27,** pp. 294–304, 1950.
3. W. J. Albersheim and J. P. Schafer, Echo Distortion in the FM Transmission of Frequency-Division Multiplex, Proc. I.R.E., **40,** pp. 316–328, 1952.

4. J. P. Vasseur, Calcul de la distortion d'une onde modulée in fréquence, Annales de Radioélectricité, **8,** pp. 20–35, 1953.
5. S. O. Rice, Mathematical Analysis of Random Noise, B.S.T.J., **23,** pp. 282–332, 1944 and **24,** pp. 46–156, 1945.
6. D. Middleton, The Distribution of Energy in Randomly Modulated Waves, Phil. Mag., **42,** pp. 689–707. July, 1951.
7. S. O. Rice, Diffraction of Plane Radio Waves by a Parabolic Cylinder, B.S.T.J., **33,** pp. 417–504, 1954.
8. E. D. Sunde, Theoretical Fundamentals of Pulse Transmission, B. S. T. J., **33,** pp. 721–788, and **33,** pp. 987–1010, May and July, 1954.

Optimum Design of Directive Antenna Arrays Subject to Random Variations

By E. N. GILBERT and S. P. MORGAN

(Manuscript received January 18, 1955)

This paper discusses the optimum design of discrete, directive antenna arrays of arbitrary geometrical configuration in space, when the excitations and spatial positions of the elements vary in a random fashion about their nominal values. Under certain assumptions the expected power pattern of an array turns out to be the power pattern of the nominal array, plus a "background" power level which has the same dependence on direction as the pattern of a single element. A set of excitations which maximizes the theoretical directivity of an array may correspond to a superdirective design, in which the background power level will completely swamp the desired pattern unless the excitations and positions of the elements are controlled with extraordinary precision. A method is given for maximizing the gain of the array while holding the expected background power level constant, when the precision with which the excitations and positions can be controlled is known. The method is illustrated with numerical examples.

1. INTRODUCTION

The effects of random variations on antenna patterns have recently been discussed in a number of papers which treat more or less special cases. Several authors[1, 2, 3] are concerned with linear Chebyshev-designed broadside arrays, and in particular with the effects of manufacturing variations on the patterns of slotted waveguide arrays. Ruze[4] has derived the expected pattern of a plane, rectangular array in which the positions of the radiators are rigidly fixed, while only their amplitudes and phases are variable. Less attention seems to have been given to the possibility that the positions of the elements may also vary in a random fashion, even though this latter situation can arise in any electromagnetic or acoustic array whose elements are not rigidly supported.

[1] L. L. Bailin and M. J. Ehrlich, I.R.E. Trans., PGAP-1, pp. 85–106, Feb., 1952.
[2] D. Ashmead, I.R.E. Trans., PGAP-4, pp. 81–92, Dec., 1952.
[3] H. F. O'Neill and L. L. Bailin, I.R.E. Trans., PGAP-4, pp. 93–102, Dec., 1952.
[4] J. Ruze, Nuovo Cimento, **9**, Supp. 3, pp. 364–380, 1952.

This paper derives a relatively simple statistical expression for the expected power pattern of an array when the excitations and positions of its elements are subject to independent random variations. The result is not restricted to linear or even to plane arrays, but is valid for arrays of arbitrary geometrical configuration in space. In the special case when the excitations are variable and the positions of the elements are either fixed or subject to random displacements with a spherically symmetric distribution, the expected power pattern is the pattern of the nominal array, plus a "background" power level which has the same dependence on direction as the pattern of a single element. The expected background power level is proportional, for small errors, to the sum of the mean-square errors in excitation and position.

We consider in particular the application of our results to the problem of designing superdirective arrays. A superdirective array is one having a beamwidth in radians much less than the reciprocal of the largest dimension of the array in wavelengths. It is well known that such narrow beams can be designed on paper, but only by employing heavy cancellation between adjacent elements. If the excitations and positions of the elements of a superdirective array are not controlled with great accuracy, the background power level due to random errors will completely swamp the desired pattern. This corresponds to the familiar fact that a really superdirective array has to be constructed with extraordinarily high precision in order to give anything like its calculated performance.

A method is given for computing the excitations which maximize the gain of an array of any specified geometrical configuration, while holding the expected background power level due to random variations constant. It is assumed that the excitation coefficients can all be controlled to the same per cent accuracy, and that the element displacements, if any, are distributed with spherical symmetry. If the random variations are taken to be zero, or if no restrictions are placed on the background power level, the present procedure becomes equivalent to the methods which have been described in recent papers[5, 6] on the maximum gain of an arbitrary array.

An interesting result of the analysis relates to arrays in which the elements are all excited with equal amplitudes, and with phases such that their fields add in phase in a specified direction in space. This arrangement may justifiably be called the normal excitation, since it is

[5] A. I. Uzkov, Dokl. Akad. Nauk SSSR, **53,** pp. 35–38, 1946.
[6] A. Bloch, R. G. Medhurst, and S. D. Pool, Proc. Inst. Elect. Engrs., **100,** Pt. III, pp. 303–314, 1953.

often adopted in practice as a means of obtaining a beam in the desired direction. It is proved that, of all possible excitations of the array which produce a main lobe in the given direction, the normal excitation leads to a pattern which is most insensitive to the effects of random errors. Furthermore, if all the elements have the same ohmic resistance, the normal excitation produces the highest power flow in the direction of the main beam for a given rate of heating the elements.

It is not invariably true that maximum gain is incompatible with minimum sensitivity of the pattern to random errors. The normal excitation maximizes the gain of any array of isotropic elements in which the distance from every element to every other element is an integral number of half wavelengths. However, for most arrays the gain can be increased above that obtainable with the normal excitation, at the expense of a (possibly enormous) increase in the sensitivity of the pattern to random errors; and some sort of compromise will have to be struck.

Several types of symmetry are commonly found in antenna arrays, and some of these symmetries force very simple relationships to hold among certain of the optimum excitation coefficients. We discuss the use of symmetry to reduce the amount of computation necessary in designing an optimum array. As an example, these considerations are applied to the design of an array of four elements located at the corners of a tetrahedron, and also to a four-element end-fire array. Curves are obtained which illustrate the relationship of gain to pattern sensitivity for several such arrays of different dimensions in wavelengths.

It is shown that if an arbitrary array of isotropic elements is excited successively to have maximum directivity in different directions, the average value of the maximum directivity over all directions in space is equal to the number of elements in the array. The excitation required to produce maximum directivity will naturally depend on direction, and considerations of pattern sensitivity are ignored. The significance of this result is that if an array configuration permits an abnormally high gain in a certain direction, as is theoretically possible, for example, with a very short end-fire array, then there exist other directions in which the maximum gain of the same array is abnormally low.

2. STATISTICAL FORMULATION

Consider an antenna array of n elements. We shall call the elements radiators, though they may equally well be thought of as receivers. We assume that each element has the same directivity pattern $\mathbf{s}(\mathbf{u})$ with

respect to a fixed set of axes,* where $\mathbf{u}$ is a unit vector representing a direction in space, and $\mathbf{s}(\mathbf{u})$ is a complex-valued vector function giving the amplitude, phase, and polarization of the radiation field over a large sphere centered at the element. For acoustic fields, $\mathbf{s}(\mathbf{u})$ is a scalar function. The average density of power flow in the direction $\mathbf{u}$ is proportional to

$$s^2(\mathbf{u}) = \mathbf{s}(\mathbf{u}) \cdot \mathbf{s}^*(\mathbf{u}) \tag{1}$$

where an asterisk denotes the complex conjugate quantity.

Let the excitation of the kth element be A_k. The complex numbers $A_1, A_2, \cdots, A_n$ will be called the excitation coefficients of the array. Let the position vector of the kth element relative to an arbitrary origin be $\mathbf{R}_k$. The field strength produced by the array at the point at the end of the vector $R\mathbf{u}$ from the origin will, for large R, be proportional to

$$\mathbf{s}(\mathbf{u}) f(\mathbf{u})/R \tag{2}$$

and the average density of power flow at the same point will be proportional to

$$\Phi(\mathbf{u})/R^2 \tag{3}$$

where the power directivity pattern $\Phi(\mathbf{u})$ is

$$\Phi(\mathbf{u}) = s^2(\mathbf{u}) \mid f(\mathbf{u}) \mid^2 \tag{4}$$

and the array factor $f(\mathbf{u})$ is

$$f(\mathbf{u}) = \sum_{k=1}^{n} A_k \exp\left(i\beta \mathbf{R}_k \cdot \mathbf{u}\right) \tag{5}$$

As usual, β denotes 2π divided by the wavelength λ.

Let us assume now that the excitation coefficients and the positions of the elements actually have some random scatter about their mean or expected values. We shall calculate the expected values of the field and power patterns. These expected values may be regarded as averages taken over a large number of different arrays, or they may be thought of as long-term time averages for a single array whose parameters vary with time in a random fashion. We can adopt the latter point of view when dealing with an array whose elements are not rigidly intercon-

* No difficulties would result in the statistical analysis from assuming a different pattern for each element, but the added generality would complicate some of the following work, and it is unnecessary for the great majority of practical arrays. We also ignore the possibility that the orientations of the elements might be subject to random variations.

nected, but are subject to displacements relative to one another in the course of time.

To make matters precise, assume that the excitation coefficients are given by

$$A_k = a_k + \alpha_k \tag{6}$$

where a_k is the expected value of A_k and the α_k's are independent random complex variables with mean zero. Also let

$$\mathbf{R}_k = \mathbf{r}_k + \boldsymbol{\varrho}_k \tag{7}$$

where $\mathbf{r}_k$ is the expected value of the position vector $\mathbf{R}_k$ and the $\boldsymbol{\varrho}_k$'s are independent random vectors with mean (0, 0, 0), all having the same statistical distribution.

We can now write down the expected values of the field and power patterns. Denoting expected values by angular brackets, we find for the field strength,

$$\begin{aligned} \langle \mathbf{s}(\mathbf{u}) f(\mathbf{u}) \rangle &= \mathbf{s}(\mathbf{u}) \sum_{k=1}^{n} \langle A_k \rangle \langle \exp\,(i\beta \mathbf{R}_k \cdot \mathbf{u}) \rangle \\ &= \mathbf{s}(\mathbf{u}) \sum_{k=1}^{n} a_k \exp\,(i\beta \mathbf{r}_k \cdot \mathbf{u}) \langle \exp\,(i\beta \boldsymbol{\varrho}_k \cdot \mathbf{u}) \rangle \\ &= \langle \exp\,(i\beta \boldsymbol{\varrho} \cdot \mathbf{u}) \rangle\, \mathbf{s}(\mathbf{u}) f_0(\mathbf{u}) \end{aligned} \tag{8}$$

where $\boldsymbol{\varrho}$ is a random vector having the same distribution as the $\boldsymbol{\varrho}_k$'s, and $f_0(\mathbf{u})$ is the nominal array factor

$$f_0(\mathbf{u}) = \sum_{k=1}^{n} a_k \exp\,(i\beta \mathbf{r}_k \cdot \mathbf{u}) \tag{9}$$

which results when the excitation coefficients and positions all have their expected values.

The norm of the general array factor may be written

$$\begin{aligned} |\, f(\mathbf{u}) \,|^2 &= \sum_{k=1}^{n} A_k \exp\,(i\beta \mathbf{R}_k \cdot \mathbf{u}) \sum_{j=1}^{n} A_j^* \exp\,(-\, i\beta \mathbf{R}_j \cdot \mathbf{u}) \\ &= \sum_{k=1}^{n} \sum_{j=1}^{n}{}' (a_k + \alpha_k)(a_j^* + \alpha_j^*) \exp\,[i\beta(\mathbf{r}_k - \mathbf{r}_j) \cdot \mathbf{u}] \exp\,[i\beta(\boldsymbol{\varrho}_k - \boldsymbol{\varrho}_j) \cdot \mathbf{u}] \\ &\qquad + \sum_{k=1}^{n} (a_k + \alpha_k)(a_k^* + \alpha_k^*) \end{aligned} \tag{10}$$

where the primed summation sign indicates that the term for which $j = k$ is to be omitted. Taking expected values and recalling that the

random variables are independent, we obtain

$$\langle | f(\mathbf{u}) |^2 \rangle$$

$$= \sum_{k=1}^{n} \sum_{j=1}^{n}{}' a_k a_j^* \exp [i\beta(\mathbf{r}_k - \mathbf{r}_j)\cdot\mathbf{u}] \langle \exp (i\beta\boldsymbol{\varrho}_k\cdot\mathbf{u}) \rangle \langle \exp (-i\beta\boldsymbol{\varrho}_j\cdot\mathbf{u}) \rangle$$

$$+ \sum_{k=1}^{n} | a_k |^2 + \sum_{k=1}^{n} \langle | \alpha_k |^2 \rangle \tag{11}$$

$$= | \langle \exp (i\beta\boldsymbol{\varrho}\cdot\mathbf{u}) \rangle |^2 | f_0(\mathbf{u}) |^2 + \sum_{k=1}^{n} \langle | \alpha_k |^2 \rangle$$

$$+ [1 - | \langle \exp (i\beta\boldsymbol{\varrho}\cdot\mathbf{u}) \rangle |^2] \sum_{k=1}^{n} | a_k |^2$$

where the last step follows by adding and subtracting the terms with $j = k$ which were omitted from the double sum.

Multiplying through by the power pattern $s^2(\mathbf{u})$ of a single element gives the expression for the expected power pattern of the array, namely

$$\langle \Phi(\mathbf{u}) \rangle = | \langle \exp (i\beta\boldsymbol{\varrho}\cdot\mathbf{u}) \rangle |^2 \Phi_0(\mathbf{u}) + s^2(\mathbf{u}) \sum_{k=1}^{n} \langle | \alpha_k |^2 \rangle$$

$$+ [1 - | \langle \exp (i\beta\boldsymbol{\varrho}\cdot\mathbf{u}) \rangle |^2] s^2(\mathbf{u}) \sum_{k=1}^{n} | a_k |^2 \tag{12}$$

where the power pattern of the nominal array is

$$\Phi_0(\mathbf{u}) = s^2(\mathbf{u}) | f_0(\mathbf{u}) |^2 \tag{13}$$

3. SPECIAL CASES

If the positions of the elements of the array are supposed to be exactly known and rigidly fixed, then the displacement vectors $\boldsymbol{\varrho}_k$ are identically zero, and the general result derived above reduces to

$$\langle \Phi(\mathbf{u}) \rangle = \Phi_0(\mathbf{u}) + s^2(\mathbf{u}) \sum_{k=1}^{n} \langle | \alpha_k |^2 \rangle \tag{14}$$

Equation (14) has a simple physical interpretation. It asserts that the expected power pattern is the power pattern of the nominal array, plus a "background" power level which has the same dependence on direction as the pattern of an individual radiator, and is proportional to the sum of the mean-square errors of the excitation coefficients. Of course for any particular array the background power will not have exactly the directional dependence $s^2(\mathbf{u})$, but will exhibit fluctuations depending on

the particular set of errors in the excitations.[7] However, in order to have the over-all pattern be a good approximation to the nominal pattern $\Phi_0(\mathbf{u})$, it is necessary to hold the expected value of the background power well below the maximum value of $\Phi_0(\mathbf{u})$. We shall discuss the implications of this requirement presently.

If the displacements of the elements are not identically zero, then we denote the cartesian components of the vector $\boldsymbol{\varrho}$ by ξ, η, ζ, and their joint probability distribution by $P(\xi, \eta, \zeta)$. $P(\xi, \eta, \zeta)\, d\xi\, d\eta\, d\zeta$ represents the probability that the end of the vector $\boldsymbol{\varrho}$, drawn from the origin, will lie in the volume element $d\xi\, d\eta\, d\zeta$ centered at (ξ, η, ζ). Then the expected value of $\exp(i\beta\boldsymbol{\varrho}\cdot\mathbf{u})$ is given by

$$\langle \exp(i\beta\boldsymbol{\varrho}\cdot\mathbf{u}) \rangle = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \exp[i\beta(\xi u_x + \eta u_y + \zeta u_z)]\, P(\xi, \eta, \zeta)\, d\xi\, d\eta\, d\zeta \tag{15}$$

where u_x, u_y, u_z are the components of the unit vector $\mathbf{u}$.

In order to evaluate the integral (15) for any particular array, one should choose the most plausible joint distribution function $P(\xi, \eta, \zeta)$ that his physical insight permits. In general $\langle \exp(i\beta\boldsymbol{\varrho}\cdot\mathbf{u}) \rangle$ will depend upon the direction of $\mathbf{u}$; but if the distribution of $\boldsymbol{\varrho}$ is spherically symmetric,* the expected value will be independent of direction. In this case $P(\xi, \eta, \zeta)$ is a function only of the magnitude ρ of the displacement vector. To evaluate the integral (15), take the ζ-axis parallel to $\mathbf{u}$, and let $p(\rho)\, d\rho$ be the probability that the length of the displacement vector lies between ρ and $\rho + d\rho$. Then

$$\begin{aligned}\langle \exp(i\beta\boldsymbol{\varrho}\cdot\mathbf{u}) \rangle &= \frac{1}{4\pi}\int_0^{2\pi}\int_0^{\pi}\int_0^{\infty} \exp(i\beta\rho\cos\theta) p(\rho)\sin\theta\, d\rho\, d\theta\, d\varphi \\ &= \int_0^{\infty} [(\sin\beta\rho)/\beta\rho]\, p(\rho)\, d\rho = \langle (\sin\beta\rho)/\beta\rho \rangle\end{aligned} \tag{16}$$

If the cartesian components of $\boldsymbol{\varrho}$ are assumed to be normally and independently distributed with mean zero and equal variances $\sigma^2/3$, so that the root-mean-square value of ρ is σ, one has

$$p(\rho) = 3(6/\pi)^{1/2}(\rho^2/\sigma^3)\exp(-3\rho^2/2\sigma^2) \tag{17}$$

and

$$\langle \exp(i\beta\boldsymbol{\varrho}\cdot\mathbf{u}) \rangle = \exp(-\beta^2\sigma^2/6) \tag{18}$$

[7] The statistical distribution of these fluctuations has been discussed by Ruze in Reference 4.

* Under some circumstances the displacement vectors may be constrained to lie in a plane. If their distribution is circularly symmetric, and if we confine our attention to directions $\mathbf{u}$ lying in the plane of the displacements, the results will be formally similar to those obtained for the spherically symmetric case.

For *any* spherically symmetric distribution of ϱ we may define a parameter δ^2 by

$$\delta^2 = | \langle \exp (i\beta\varrho \cdot \mathbf{u}) \rangle |^{-2} - 1 \tag{19}$$

If σ, the root-mean-square value of ρ, is small compared to the wavelength, then

$$\delta^2 \approx \beta^2\sigma^2/3 \tag{20}$$

for the normal distribution (17), and also for other distributions which taper off for large ρ with comparable or greater rapidity.

From equations (12) and (19) we obtain the normalized expected power pattern for a spherically symmetric distribution of displacements, namely

$$(1 + \delta^2) \langle \Phi(\mathbf{u}) \rangle = \Phi_0(\mathbf{u}) + s^2(\mathbf{u}) \left[(1 + \delta^2) \sum_{k=1}^{n} \langle | \alpha_k |^2 \rangle + \delta^2 \sum_{k=1}^{n} | a_k |^2 \right] \tag{21}$$

Again the expected pattern turns out to be the nominal pattern plus a background level with the same distribution as the pattern of a single element.

In what follows we shall idealize the problem somewhat by assuming that the excitation coefficients A_k can all be controlled to the same relative accuracy.[8] Precisely we suppose there is a small number ϵ such that

$$\langle |\alpha_k|^2 \rangle = \epsilon^2 | a_k |^2, \qquad k = 1, 2, \cdots, n \tag{22}$$

Thus (21) becomes

$$(1 + \delta^2) \langle \Phi(\mathbf{u}) \rangle = \Phi_0(\mathbf{u}) + \Delta^2 s^2(\mathbf{u}) \sum_{k=1}^{n} | a_k |^2 \tag{23}$$

where

$$\Delta^2 = (1 + \delta^2)\epsilon^2 + \delta^2 \approx \delta^2 + \epsilon^2 \tag{24}$$

and the last approximation is valid if δ^2 is small compared to unity.

[8] Ruze's assumptions in Reference 4 amount to taking $| A_k |$ and arg A_k as independent random variables with means I_k, ϑ_k, where the coefficients of the desired pattern are $I_k \exp(i\vartheta_k)$. All the phases are assumed to be normally distributed with the same variance, and the variance of the amplitude $| A_k |$ is taken proportional to $| I_k |^2$. From these assumptions equation (22) follows. Ruze's result looks more complicated than equation (14) mainly because the expected value of A_k differs from $I_k \exp(i\vartheta_k)$ by a constant factor.

4. CRITERIA FOR GOOD PATTERNS

The preceding statistical analysis is equally valid whether the nominal pattern $\Phi_0(\mathbf{u})$ is to have some specified beam shape for a particular application, or merely to provide as narrow a beam as possible. However the latter case is of greater practical interest, and so we shall consider henceforth only the design of highly directive or pencil beam arrays.

We suppose that the number and configuration of the radiators are fixed in advance, and we have to choose the excitation coefficients to obtain the desired pattern of the nominal array. We shall denote the set of n complex numbers $(a_1, a_2, \cdots, a_n)$ by the single symbol a. Our problem is to find a way of exciting the array so as to produce a narrow beam in some direction $\mathbf{u}_0$, i.e., we must choose a in such a way as to make $\Phi_0(\mathbf{u}_0)$ small for all directions $\mathbf{u}$ not near $\mathbf{u}_0$. When we try to formulate the problem more precisely than this we find a large number of alternatives.

As a mathematically tractable criterion for a good pattern, we shall stipulate that a is to be chosen so as to minimize the generalized gain function

$$G(a) = \frac{\Phi_0(\mathbf{u}_0)}{\frac{1}{4\pi}\int_S \Phi_0(\mathbf{u})w(\mathbf{u})\, d\Omega} \tag{25}$$

where $w(\mathbf{u})$ is a non-negative weight function which may be chosen at pleasure, $d\Omega$ is an element of solid angle, and S is the surface of the unit sphere. Since $\Phi_0(\mathbf{u})$ is proportional to the density of power flow in the direction $\mathbf{u}$, $G(a)$ represents the ratio of power flow in the direction $\mathbf{u}_0$ to a weighted average of power flow in all directions over a sphere. If one were interested in the radiation pattern only in some plane containing $\mathbf{u}_0$, the weighted average could be taken in all directions around a circle with only minor changes in the formal analysis.

A few comments may clarify the significance of the function $w(\mathbf{u})$. If $w(\mathbf{u})$ is taken to be identically unity, then $G(a)$ is just the conventional gain of an array with the set of excitations a. However, it is well known that merely maximizing the gain does not always produce a pattern with low side lobes. If it is important to prevent large fields from being radiated in specified directions, one may choose $w(\mathbf{u})$ to be unity over the set of unwanted directions and zero elsewhere; and in principle even more complicated choices of $w(\mathbf{u})$ could be made to discourage side lobes. For receiving arrays, $w(\mathbf{u})$ may similarly be chosen to discriminate against the reception of spurious signals from particular directions.

In terms of the excitation coefficients, the function $G(a)$ has a simple formal expression, namely

$$G(a) = s^2(\mathbf{u}_0) \frac{\left| \sum_{k=1}^{n} a_k \exp(i\beta \mathbf{r}_k \cdot \mathbf{u}_0) \right|^2}{\sum_{k=1}^{n} \sum_{j=1}^{n} h_{jk} a_k a_j^*} \tag{26}$$

where

$$h_{jk} = \frac{1}{4\pi} \int_S \exp[i\beta(\mathbf{r}_k - \mathbf{r}_j) \cdot \mathbf{u}] s^2(\mathbf{u}) w(\mathbf{u})\, d\Omega \tag{27}$$

and obviously

$$h_{kj} = h_{jk}^* \tag{28}$$

The coefficients h_{jk} depend only on the weight function, the pattern of a single element, and the positions of the elements (in terms of wavelength). When the elements are isotropic and the weight function is identically unity, we have

$$h_{jk} = \frac{\sin \beta r_{jk}}{\beta r_{jk}} \tag{29}$$

where r_{jk} is the distance from radiator j to radiator k. In any case $G(a)$ is the quotient of two Hermitian forms in the excitation coefficients, and can therefore be maximized by standard mathematical techniques.

It may be noted that the use of an integrated criterion such as (25) for keeping the fields small away from the main beam does not absolutely guarantee the absence of undesirably high side lobes in particular directions for any given array. To be sure of keeping all side lobes below a certain level, we should choose a to maximize some such expression as

$$T(a) = \frac{\Phi_0(\mathbf{u}_0)}{\max_{\Omega} \Phi_0(\mathbf{u})} \tag{30}$$

where the denominator is the maximum value of $\Phi_0(\mathbf{u})$ over a chosen set of directions Ω not containing $\mathbf{u}_0$. For equispaced linear arrays the criterion (30) leads to the Chebyshev design procedure first described by Dolph. However in general it is a much harder mathematical problem to maximize $T(a)$ than to maximize $G(a)$, and for that reason we shall not employ $T(a)$ here.

5. RESTRICTIONS ON SUPERDIRECTIVE ARRAYS

Even when a set of excitation coefficients maximizing the gain function $G(a)$ has been found, there is no assurance that this set will be a satis-

factory one on which to base the construction of a physical array. There is a further restriction on the solution: the array must be excited in such a way that when it is constructed it is likely to have a pattern $\Phi(\mathbf{u})$ which differs from the nominal pattern $\Phi_0(\mathbf{u})$ by an acceptably small amount.

From Section 3 the expected power pattern is of the form

$$\Phi_0(\mathbf{u}) + \Delta^2 s^2(\mathbf{u}) \sum_{k=1}^{n} | a_k |^2 \tag{31}$$

where Δ^2 includes the effects of both excitation and position errors. The background power level relative to the main lobe of the nominal pattern is then just

$$\frac{\Delta^2 s^2(\mathbf{u}) \sum_{k=1}^{n} | a_k |^2}{\Phi_0(\mathbf{u}_0)} = \frac{\Delta^2 s^2(\mathbf{u})}{s^2(\mathbf{u}_0)} K(a) \tag{32}$$

where

$$K(a) = \frac{\sum_{k=1}^{n} | a_k |^2}{\left| \sum_{k=1}^{n} a_k \exp (i\beta \mathbf{r}_k \cdot \mathbf{u}_0) \right|^2} \tag{33}$$

$K(a)$ is a function measuring the susceptibility of the pattern to random errors in the excitations and positions of the elements. Since in practice Δ^2 is never zero, an array with too large a value of $K(a)$ will be unacceptable.

Although the function $K(a)$ has been introduced as a result of statistical considerations, it can also be interpreted in terms of the efficiency of the array as an energy radiator. If we imagine the elements to have a certain ohmic resistance, and the excitation coefficients to correspond to the element currents, then $\Sigma | a_k |^2$ is a measure of the power which is lost in the form of heat, and $K(a)$ is proportional to the ratio of dissipated power to power density in the direction $\mathbf{u}_0$. Thus a large value of $K(a)$ corresponds to large circulating currents in the array, and to high ohmic losses for a given rate of radiation of power in the direction of the main beam.

As a simple example of arrays which can have arbitrarily high gain at the expense of large values of $K(a)$, let us consider an end-fire array of length L pointing in the direction $\mathbf{u}_0$ (say the direction of the z-axis). The array will have $n + 1$ elements situated at $z = 0, L/n, 2L/n, \cdots, L$. Let the expected excitation coefficient of the radiator at $z = kL/n$ be

$$a_k = (-)^k C_{n,k} \exp (ik\beta L/n) \tag{34}$$

where the $C_{n,k}$'s are binomial coefficients. Then from (9) the norm of the nominal array factor is

$$\begin{aligned} |f_0(\mathbf{u})|^2 &= |1 - \exp[i(\mathbf{u}_0\cdot\mathbf{u} + 1)\beta L/n]|^{2n} \\ &= 2^{2n}\sin^{2n}[(1+\cos\theta)\beta L/2n] \\ &= 2^{2n}\sin^{2n}[(\beta L\cos^2 \tfrac{1}{2}\theta)/n] \end{aligned} \tag{35}$$

where θ is the angle between the direction $\mathbf{u}$ and the direction of the array. If L is fixed and n is large, then approximately

$$|f_0(\mathbf{u})|^2 \approx (2\beta L/n)^{2n}\cos^{4n}\tfrac{1}{2}\theta. \tag{36}$$

Taking n large enough, one obtains an arbitrarily sharp beam and an arbitrarily high gain. On the other hand,

$$\sum_{k=1}^{n}|a_k|^2 = \sum_{k=1}^{n}C_{n,k}^{\ 2} = \frac{(2n)!}{(n!)^2} \approx \frac{2^{2n}}{(n\pi)^{1/2}} \tag{37}$$

the last approximation being valid for large n. Hence for L fixed and n large, we have from (33),

$$K(a) \approx (n/\beta L)^{2n}(n\pi)^{-1/2} \tag{38}$$

If Δ is a typical figure like 0.01, it is clear that this end-fire array is totally useless, in spite of its high theoretical gain, if the spacing L/n is much less than $1/\beta = \lambda/2\pi$.

The array just considered exhibits a characteristic feature of superdirective arrays. All such arrays depend for their narrow beamwidths on heavy cancellation between the fields of closely spaced radiators. In practice the radiators can be adjusted with only finite precision, and the random errors which contribute to the background power level will not cancel out, on the average, as do the fields of the nominal elements in all directions except that of the main beam. Hence the background power level will completely swamp the main beam unless the array is designed with extraordinarily high precision.

It turns out that merely maximizing the gain function $G(a)$ of an array with interelement spacings of less than about a quarter wavelength is likely to lead to a superdirective design with an unacceptably large value of $K(a)$. To get useful results, therefore, we should maximize $G(a)$ subject to the auxiliary condition that $K(a)$ is not to exceed a preassigned value. A method for doing this is given in the next section.

We note that it would appear possible on paper to design an array in a limited volume with high $G(a)$ and low $K(a)$ by using a large number of closely packed elements. For example, suppose we have an array with

a high value of $G(a)$. Now suppose each element to be divided into m very nearby elements, each with $1/m$ times the original excitation. The nominal field pattern, which determines the theoretical gain, is essentially unchanged, while $K(a)$ is divided by m. The catch is that a really superdirective array with a reasonable value of Δ^2 would require a colossal number of elements to reduce $K(a)$ to an acceptable value. Furthermore our statistical arguments have been based on the assumption that the excitation coefficients may be regarded as *independent* random variables when the antenna is built. If the elements are packed close together it seems unlikely that the excitations remain independent; then the use of $K(a)$ to determine the precision which must be maintained is no longer justified.

In dealing with superdirective arrays it should always be remembered that a superdirective antenna is a high-Q, small-bandwidth device, inasmuch as the amount of reactive energy in the near field of the antenna is very large compared to the energy radiated per cycle. The resulting stringent physical limitations on superdirectivity have been exhibited by Chu.[9] We shall not, however, discuss questions of bandwidth here.

6. OPTIMUM DESIGNS TAKING ACCOUNT OF FINITE PRECISION

The procedure for maximizing the gain function $G(a)$ of a definite array, while requiring $K(a)$ not to exceed a specified value, depends upon certain theorems which will now be stated. The proofs are given in the appendix.

Theorem I. Let $\mu \geqq 0$ and let $a[\mu]$ denote the set of excitation coefficients satisfying the system of linear equations

$$\sum_{j=1}^{n} h_{kj} a_j[\mu] + \mu a_k[\mu] = \exp(-i\beta \mathbf{r}_k \cdot \mathbf{u}_0) \tag{39}$$

$k = 1, 2, \cdots, n$. Of all possible choices of a satisfying

$$K(a) \leqq K(a[\mu]) \tag{40}$$

the maximum value of $G(a)$ is obtained when $a = a[\mu]$.

The parameter μ is essentially a Lagrangian undetermined multiplier, such as is commonly used in the calculus of variations and in the determination of maxima or minima subject to a constraint. When designing a directive array we may select a reasonable value of $K(a)$, say K_0, with an eye to the precision with which we expect to construct the array.

[9] L. J. Chu, J. App. Phys., **19**, pp. 1163–1175, 1948.

What we should then like to do is choose the value of μ for which

$$K(a[\mu]) = K_0 \tag{41}$$

and solve (39) with this value of μ. Unfortunately we cannot determine μ simply or directly from the condition (41), so we may have to make several trials with different values of μ. If our first guess does not yield a value of K sufficiently close to K_0, the direction to proceed is indicated by

Theorem II. Both $K(a[\mu])$ and $G(a[\mu])$ are monotone nonincreasing functions of μ; that is, $dK(a[\mu])/d\mu \leqq 0$ and $dG(a[\mu])/d\mu \leqq 0$.

Fairly simple expressions may be obtained (see equations (A4), (A5), and (A20) through (A24) of the appendix) for $a[\mu]$, $K(a[\mu])$, and $G(a[\mu])$ in terms of the eigenvalues and eigenvectors of the Hermitian matrix (h_{jk}), and the parameter μ. To make a thorough study of any particular array configuration, one might well start by computing the eigenvalues and eigenvectors. Then it would be easy to plot K and G against μ over any desired range.

A restriction on the possible values of $K(a)$ is given by

Theorem III. For any array with n elements,

$$K(a) \geqq 1/n \tag{42}$$

and the coefficients which yield the value $1/n$ are

$$a_k = \exp(-i\beta \mathbf{r}_k \cdot \mathbf{u}_0), \quad k = 1, 2, \cdots, n \tag{43}$$

up to a constant proportionality factor.

The choice of coefficients (43) means that all elements have equal amplitudes and the phases are chosen to make their contributions add in phase in the direction $\mathbf{u}_0$. This may be called the normal excitation; it is often adopted in practice as a means of obtaining a beam in the direction $\mathbf{u}_0$. The patterns so obtained are most insensitive to random errors in the excitation coefficients, although frequently at the expense of rather disagreeable side lobes. Since, as pointed out in Section 5, $K(a)$ also measures the efficiency of the array as an energy radiator, Theorem III shows that the normal excitation produces the highest power flow in the direction $\mathbf{u}_0$ for a given rate of heating the elements.

If no auxiliary condition is imposed on the value of $K(a)$, we have

Theorem IV. The coefficients which maximize the gain function $G(a)$ absolutely are obtained by putting $\mu = 0$ in equations (39), so that $a = a[0]$, where

$$\sum_{j=1}^{n} h_{kj} a_j[0] = \exp(-i\beta \mathbf{r}_k \cdot \mathbf{u}_0) \tag{44}$$

$k = 1, 2, \cdots, n$.

Equations (44) are equivalent to those given by Bloch, Medhurst, and Pool[10] as a result of a differently phrased argument. They could also be obtained by following up Uzkov's remark[11] that the problem of maximizing the directivity of an arbitrary array amounts to the problem of choosing an orthogonal basis for a complex linear vector space generated by the patterns of the individual radiators. In principle both these papers show how to determine the maximum gain in the general case when each element of the array has a different directivity pattern; but they say nothing about precision requirements.

It is not invariably true that maximum gain is incompatible with minimum sensitivity to random errors. For example, the normal excitation of Theorem III, which minimizes $K(a)$, simultaneously maximizes the conventional gain of any array of isotropic elements in which the distance from every element to every other element is an integral number of half wavelengths. Under this condition (29) shows that $h_{jk} = \delta_{jk}$; and the values of G and K obtained from the solution of (39) are independent of μ. An example is a linear array with half-wavelength spacing between adjacent elements; another is the three-dimensional configuration of four elements at the vertices of a tetrahedron of edge $\lambda/2$. If the elements of these arrays are not isotropic, then the normal excitation does not give quite the maximum gain; but the difference will be small if the beamwidth of the array factor is narrow compared to the beamwidth of the element pattern. For a general array with interelement spacings much less than a half wavelength, however, there may be a great deal of difference between the excitation and pattern for minimum sensitivity ($\mu = \infty$) and the excitation and pattern for maximum gain ($\mu = 0$). Illustrative examples are worked out in Section 8.

7. SYMMETRIC ARRAYS

To find optimum excitation coefficients for an n-element array by the method of the preceding section requires the solution of a system of n simultaneous linear equations, and this can be laborious if n is large. Fortunately several types of symmetry are commonly found in antenna arrays. Some of these symmetries force very simple relationships to hold among certain of the optimum excitation coefficients, and can therefore be used to obtain an immediate reduction of the order of the system of equations (39).

By a *symmetry* S of an antenna configuration we shall mean any combination of translations, rotations, and reflections which leaves the con-

[10] Reference 6, p. 304, equation (6b).
[11] Reference 5, p. 37.

figuration invariant. Let Sk denote the number of the element of the array into which the kth element is carried by the symmetry S. Some simple symmetries and the corresponding relationships which can exist between a_k and a_{Sk} are listed in Table I. These relationships hold for the solutions $a_k[\mu]$ of equations (39) if the origin of the coordinate system is symmetrically placed,* i.e., if the origin is invariant under the symmetry operation S. In the last four entries of the table we assume that the pattern $s^2(\mathbf{u})$ of a single element and the weight function $w(\mathbf{u})$ appearing in the definition of the generalized gain depend only on the angle between $\mathbf{u}$ and $\mathbf{u}_0$.

To illustrate the method of proving these relationships consider the second one. Reflection in a plane normal to $\mathbf{u}_0$ carries a point $\mathbf{r}$ into

$$S\mathbf{r} = \mathbf{r} - 2(\mathbf{r}\cdot\mathbf{u}_0)\mathbf{u}_0 \tag{45}$$

and for any points $\mathbf{x}$ and $\mathbf{y}$

$$\mathbf{u}_0\cdot S\mathbf{x} = -\mathbf{u}_0\cdot\mathbf{x} \tag{46}$$

$$S\mathbf{x}\cdot S\mathbf{y} = \mathbf{x}\cdot\mathbf{y} \tag{47}$$

From (46) it follows that

$$s^2(-S\mathbf{u})w(-S\mathbf{u}) = s^2(\mathbf{u})w(\mathbf{u}) \tag{48}$$

since by assumption $s^2(\mathbf{u})$ and $w(\mathbf{u})$ depend only on the angle between $\mathbf{u}$ and $\mathbf{u}_0$, that is, on $\cos^{-1}(\mathbf{u}\cdot\mathbf{u}_0)$. From the last three equations and (27) we conclude that

$$\begin{aligned} h_{Sk,Sj} &= \frac{1}{4\pi}\int_S \exp\,[-i\beta(\mathbf{r}_{Sj} - \mathbf{r}_{Sk})\cdot S\mathbf{u}]s^2(-S\mathbf{u})w(-S\mathbf{u})\,d\Omega \\ &= \frac{1}{4\pi}\int_S \exp\,[-i\beta(\mathbf{r}_j - \mathbf{r}_k)\cdot\mathbf{u}]s^2(\mathbf{u})w(\mathbf{u})\,d\Omega \\ &= h_{kj}^* \end{aligned} \tag{49}$$

We also have

$$\exp\,(-i\beta\mathbf{r}_{Sk}\cdot\mathbf{u}_0) = [\exp\,(-i\beta\mathbf{r}_k\cdot\mathbf{u}_0)]^* \tag{50}$$

The system of equations (39) can be written in the form

$$\sum_{j=1}^{n} h_{Sk,Sj}a_{Sj} + \mu a_{Sk} = \exp\,(-i\beta\mathbf{r}_{Sk}\cdot\mathbf{u}_0) \tag{51}$$

* If the origin is not symmetrically placed it may be necessary to multiply $a[\mu]$ by a suitably chosen complex constant in order to get a set of excitation coefficients which satisfy the relationships of Table I.

or, using (49) and (50),

$$\sum_{j=1}^{n} h_{kj}{}^{*} a_{Sj} + \mu a_{Sk} = [\exp(-i\beta \mathbf{r}_k \cdot \mathbf{u}_0)]^{*} \tag{52}$$

or finally, taking complex conjugates,

$$\sum_{j=1}^{n} h_{kj} a_{Sj}{}^{*} + \mu a_{Sk}{}^{*} = \exp(-i\beta \mathbf{r}_k \cdot \mathbf{u}_0) \tag{53}$$

Comparing (39) and (53) we see that a_k and $a_{Sk}{}^{*}$ are solutions of the same system of linear equations. The determinant of the system is not zero, by equation (A7) of the appendix, and hence

$$a_k = a_{Sk}{}^{*} \tag{54}$$

Proofs for the other cases in Table I are similar.

In cases 3 and 5 of Table I the relationship $a_{Sk} = a_k$ reduces the system of equations (39) to a system of order equal to the number of different classes of symmetric points in the array. In cases 1, 2, and 4 the relationship $a_{Sk} = a_k{}^{*}$ is more difficult to exploit. If h_{jk} is a real matrix, which will be true whenever

$$s^2(-\mathbf{u})w(-\mathbf{u}) = s^2(\mathbf{u})w(\mathbf{u}) \tag{55}$$

the system (39) may be rewritten as the following pair of real systems:

$$\begin{aligned} \sum_{j=1}^{n} h_{kj} \operatorname{Re} a_j + \mu \operatorname{Re} a_k &= \cos(\beta \mathbf{r}_k \cdot \mathbf{u}_0) \\ \sum_{j=1}^{n} h_{kj} \operatorname{Im} a_j + \mu \operatorname{Im} a_k &= -\sin(\beta \mathbf{r}_k \cdot \mathbf{u}_0) \end{aligned} \tag{56}$$

Now using

$$\operatorname{Re} a_{Sj} = \operatorname{Re} a_j, \qquad \operatorname{Im} a_{Sj} = -\operatorname{Im} a_j \tag{57}$$

the order of these systems reduces to about $n/2$ each.

TABLE I — RELATIONSHIPS AMONG OPTIMUM EXCITATION COEFFICIENTS IN SYMMETRIC ANTENNA ARRAYS.

Case	Symmetry S	Relationship
1	Reflection in a point	$a_{Sk} = a_k{}^{*}$
2	Reflection in a plane normal to $\mathbf{u}_0$	$a_{Sk} = a_k{}^{*}$
3	Reflection in a plane parallel to $\mathbf{u}_0$	$a_{Sk} = a_k$
4	180° rotation about an axis perpendicular to $\mathbf{u}_0$	$a_{Sk} = a_k{}^{*}$
5	Rotation about an axis parallel to $\mathbf{u}_0$	$a_{Sk} = a_k$

It may be noted that the relationships listed in Table I still apply if instead of $G(a)$ the function to be maximized is the main-beam to side-lobe ratio mentioned at the end of Section 4, namely

$$T(a) = \frac{\Phi_0(\mathbf{u}_0)}{\max_{\Omega} \Phi_0(\mathbf{u})} \tag{58}$$

subject to the auxiliary condition on $K(a)$. In this case we assume that the set Ω of directions in which radiation is to be kept small has rotational symmetry about the direction $\mathbf{u}_0$, that is, if $\mathbf{u}_1$ is in Ω any other $\mathbf{u}$ for which $\mathbf{u}\cdot\mathbf{u}_0 = \mathbf{u}_1\cdot\mathbf{u}_0$ is also in Ω.

8. NUMERICAL EXAMPLES

We shall apply the preceding theory to two array configurations which illustrate some points of interest without requiring unnecessarily heavy computation. For simplicity throughout this section we consider only isotropic elements and adopt the conventional definition of gain, so that the matrix elements h_{jk} are of the form given by equation (29), namely $(\sin \beta r_{jk})/\beta r_{jk}$.

The first example consists of four elements situated at the vertices of a tetrahedron of edge length l (Fig. 1). The unit vector $\mathbf{u}_0$ is taken to be along the altitude of the tetrahedron which passes through the element a_4 and is perpendicular to the plane containing a_1, a_2, and a_3. This array is invariant under rotations of 120° about an axis containing $\mathbf{u}_0$, and so by Case 5 of Table I we have

$$a_1 = a_2 = a_3 \tag{59}$$

It is easily shown that the system of equations (39) for the optimum coefficients reduces to the two equations:

$$\begin{aligned} \left[1 + \mu + \frac{2 \sin \beta l}{\beta l}\right] a_1 + \frac{\sin \beta l}{\beta l} a_4 &= 1 \\ \frac{3 \sin \beta l}{\beta l} a_1 + (1 + \mu) a_4 &= \exp\left[-i(\tfrac{2}{3})^{1/2} \beta l\right] \end{aligned} \tag{60}$$

After solving (60), the functions $G(a)$ and $K(a)$ may be computed from their definitions (26) and (33).

The optimum gain G is plotted against the sensitivity function K in Fig. 1 for $l = \lambda/2$, $\lambda/4$, $\lambda/8$, and $\lambda/16$. As a matter of interest a few values of the parameter μ are shown on the curves. The curve for $l = \lambda/2$ consists of the single point $K = \frac{1}{4}$, $G = 4$. This is an array with half-

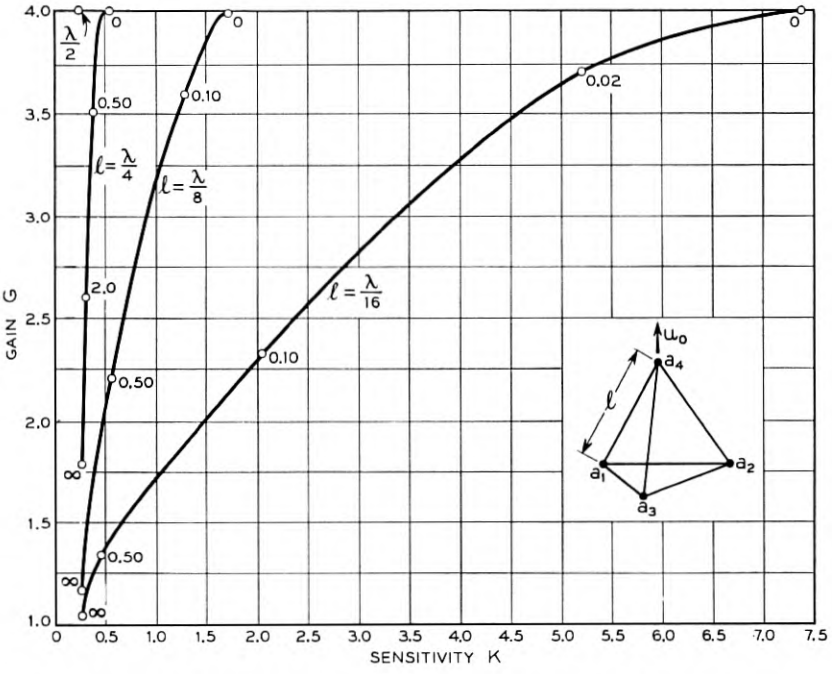


Fig. 1 — Gain G versus sensitivity K for optimum tetrahedral arrays of various edge lengths l. Some values of the parameter μ are shown on the curves.

wavelength interelement spacing, for which, as noted at the end of Section 6, minimum sensitivity and maximum gain are obtained simultaneously. The lower ends of the curves ($\mu = \infty$) all represent normal excitation and minimum sensitivity; in accord with Theorem III the limiting value of K is $\frac{1}{4}$. The corresponding gain decreases as the size of the array decreases. As the condition on K is relaxed by decreasing the parameter μ, the gain increases to a value very near 4 in all cases, and the sensitivity increases to larger and larger values for the smaller arrays. The maximum values of G, which occur for $\mu = 0$, and the corresponding values of K are listed in Table II.

As a second example we consider a four-element equispaced linear end-fire array with interelement spacing l (Fig. 2). This array is symmetric with respect to its center point, and so by Case 1 of Table I,

$$\begin{aligned} a_4 &= a_1^* \\ a_3 &= a_2^* \end{aligned} \tag{61}$$

Equations (39) break up into two pairs of real equations of the form

TABLE II — MAXIMUM GAIN G AND CORRESPONDING SENSITIVITY K FOR FOUR-ELEMENT TETRAHEDRAL AND LINEAR ARRAYS WITH INTERELEMENT SPACINGS l/λ WAVELENGTHS.

l/λ	Tetrahedral		Linear	
	G	K	G	K
1/2	4.000	0.2500	4.000	0.2500
1/4	3.960	0.5405	12.77	2.065
1/8	3.990	1.901	15.21	1.07×10^2
1/16	3.997	7.371	15.80	6.6×10^3
1/32	—	—	15.95	4.2×10^5
0	4.000	∞	16.00	∞

(56), namely:

$$\left[1 + \mu + \frac{\sin 3\beta l}{3\beta l}\right] \text{Re } a_1 + \left[\frac{\sin \beta l}{\beta l} + \frac{\sin 2\beta l}{2\beta l}\right] \text{Re } a_2 = \cos 3\beta l/2$$
$$\left[\frac{\sin \beta l}{\beta l} + \frac{\sin 2\beta l}{2\beta l}\right] \text{Re } a_1 + \left[1 + \mu + \frac{\sin \beta l}{\beta l}\right] \text{Re } a_2 = \cos \beta l/2 \tag{62}$$

and

$$\left[1 + \mu - \frac{\sin 3\beta l}{3\beta l}\right] \text{Im } a_1 + \left[\frac{\sin \beta l}{\beta l} - \frac{\sin 2\beta l}{2\beta l}\right] \text{Im } a_2 = -\sin 3\beta l/2$$
$$\left[\frac{\sin \beta l}{\beta l} - \frac{\sin 2\beta l}{2\beta l}\right] \text{Im } a_1 + \left[1 + \mu - \frac{\sin \beta l}{\beta l}\right] \text{Im } a_2 = -\sin \beta l/2 \tag{63}$$

Plots of G against K, as computed from the above equations, are shown in Fig. 2 for arrays with $l = \lambda/4$, $\lambda/8$, $\lambda/16$, and $\lambda/32$. Note that the scale of Fig. 2 is different from the scale of Fig. 1. The point for $l = \lambda/2$, which again falls at $K = \frac{1}{4}$, $G = 4$, is not shown because it happens to coincide with the lower end of the curve for $l = \lambda/4$. For spacings less than $\lambda/2$ the curves rise to values of gain greater than 4, the limiting value of gain for infinitesimal spacing being 16. However, for spacings less than $\lambda/4$ the sensitivity becomes very large for the higher gains, so that the upper ends of the curves are far off the horizontal scale in Fig. 2. The theoretical maximum gains and corresponding sensitivities are listed in Table II.

A considerable difference is evident in the behavior of small tetrahedral and small linear arrays, where by "small" we mean interelement spacings appreciably less than a half wavelength.* Sensitivity considerations

* It is probably not meaningful to make a direct comparison between tetrahedral and linear arrays of the same spacing l, on account of their very different proportions. For example, the tetrahedral array will fit into a sphere of diameter $1.225l$, while the linear array requires a sphere of diameter $3l$.

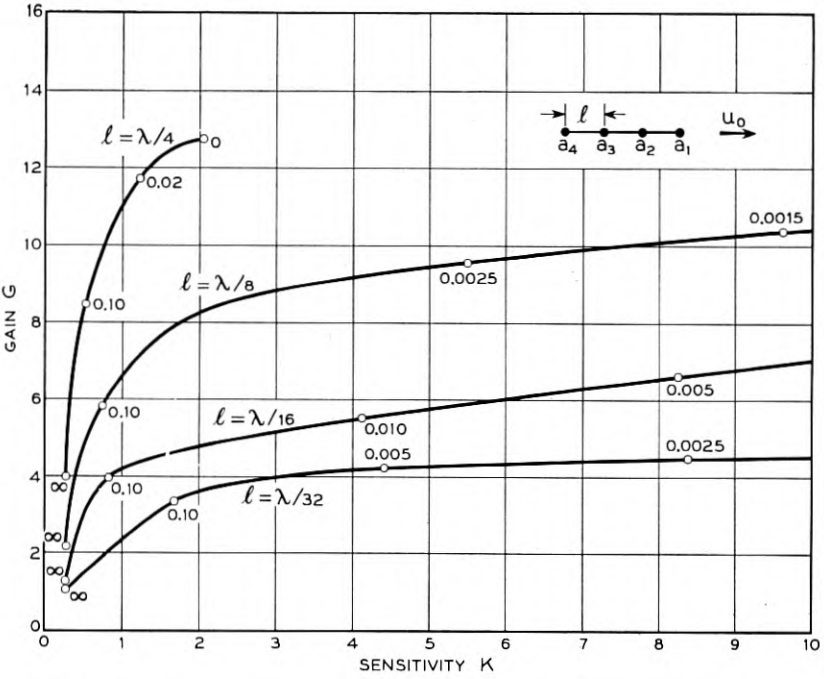


Fig. 2 — Gain G versus sensitivity K for optimum four-element linear end-fire arrays with various interelement spacings l. Some values of the parameter μ are shown on the curves.

being ignored for the moment, the small linear array can have much higher gains *in the end-fire direction* than the tetrahedral array can have in any direction. On the other hand, by a suitable choice of excitation the tetrahedral array can exhibit very nearly the same gain in an arbitrary direction, while in directions different from end-fire the small linear array cannot produce nearly so high a gain. These observations serve to introduce a more general result, as follows:

Let $\Gamma(\mathbf{u})$ be the maximum (conventional) gain which can be achieved in the direction $\mathbf{u}$ with an array of n isotropic elements, located at arbitrary preassigned positions in space. The value of $\Gamma(\mathbf{u})$ and the excitation necessary to obtain it will of course depend upon $\mathbf{u}$. The following theorem is proved in the appendix:

Theorem V. The average value over all directions in space of the maximum gain $\Gamma(\mathbf{u})$ *of an array of isotropic elements is equal to the number of elements in the array; that is,*

$$\frac{1}{4\pi}\int_S \Gamma(\mathbf{u})\, d\Omega = n \tag{64}$$

From this theorem it is obvious that the tetrahedral array, which comes as close to spherical symmetry as is possible with four elements, will have a maximum gain of approximately 4 in every direction, while the small linear array, which has a maximum gain of nearly 16 in the end-fire direction, must have a maximum gain much smaller than 4 in most other directions. In the limiting case of an infinitesimally short linear array of n elements, it may be shown by a simple extension of Uzkov's derivation[12] that the maximum gain which can be obtained in the direction $\mathbf{u}$ is

$$\Gamma(\mathbf{u}) = \sum_{k=0}^{n-1} (2k + 1)P_k^2 (\cos \theta) \tag{65}$$

where $P_k(\cos \theta)$ is a Legendre polynomial and θ is the angle between $\mathbf{u}$ and the direction of the array. $\Gamma(\mathbf{u})$ is equal to n^2 when $\theta = 0$, but its value decreases rapidly as θ varies away from zero. It is easy to verify Theorem V in this special case, since we have

$$\begin{aligned} \frac{1}{4\pi} \int_S \Gamma(\mathbf{u})\, d\Omega &= \frac{1}{2} \sum_{k=0}^{n-1} (2k + 1) \int_0^{\pi} P_k^2 (\cos \theta) \sin \theta\, d\theta \\ &= \frac{1}{2} \sum_{k=0}^{n-1} (2k + 1)[2/(2k + 1)] = n \end{aligned} \tag{66}$$

An infinitesimal array is not in itself of practical importance, since for gains greater than unity the excitation coefficients and the sensitivity are infinite; but the above example does illustrate that if an array can have an abnormally high gain in certain directions, there must be other directions in which its maximum gain is abnormally low. This fact may have applications to the design of arrays whose physical orientation is fixed, while the main beam is caused to scan by electrically varying the excitations of the various elements.

In conclusion we should like to point out that the design procedure given in this paper applies to any array in which the number of elements and their configuration are specified in advance. In many practical cases the array to be used will be so specified, and our method then yields the optimum design. One could, of course, ask a more general question: Given n elements and a sphere of fixed diameter, how should the elements be placed inside the sphere and how should they be excited to maximize the gain in a specified direction, while holding the function $K(a)$, which measures the sensitivity of the pattern to random variations, below a preassigned value? The restriction to a definite number of

[12] Reference 5, pp. 37–38.

elements is essential, since as pointed out in Section V we could produce on paper an arbitrarily high gain with an arbitrarily small $K(a)$ by packing colossal numbers of elements into the sphere. It may well be true (though a proof seems to be lacking) that if there is no restriction on pattern sensitivity, the maximum gain possible with n elements is obtained by arranging them in an end-fire array of infinitesimal length; but the excitation coefficients and $K(a)$ are then infinite. It is not self-evident that the end-fire arrangement will be optimum if $K(a)$ is required to have a reasonable, finite value. The very interesting problem of determining optimum array configurations for finite values of $K(a)$ appears, however, to be considerably more difficult than the problem of determining optimum excitations for arrays of fixed configuration.

Appendix

Proofs of optimization theorems

Let H denote the Hermitian matrix (h_{jk}), a the n-vector whose components are the excitation coefficients, and e the n-vector whose components are

$$e_k = \exp(-i\beta \mathbf{r}_k \cdot \mathbf{u}_0), \qquad k = 1, 2, \cdots, n \tag{A1}$$

The product Ha of the matrix H and the vector a is the vector whose jth component is

$$(Ha)_j = \sum_{k=1}^{n} h_{jk} a_k, \quad j = 1, 2, \cdots, n \tag{A2}$$

and the inner product (x, y) of two vectors is defined as

$$(x, y) = \sum_{k=1}^{n} x_k y_k^* \tag{A3}$$

The functions $G(a)$ and $K(a)$ defined by (26) and (33) may accordingly be written

$$G(a) = \frac{|(a, e)|^2}{(Ha, a)} \tag{A4}$$

$$K(a) = \frac{(a, a)}{|(a, e)|^2} \tag{A5}$$

Theorem I. Consider the system of equations

$$Ha + \mu a = e, \tag{A6}$$

where μ is a scalar parameter. If $\mu \geqq 0$, then for any nonzero a whatever,

$$(Ha, a) + \mu(a, a) = \frac{1}{4\pi} \int_S |f_0(\mathbf{u})|^2 s^2(\mathbf{u}) w(\mathbf{u})\, d\Omega + \mu(a, a) > 0 \tag{A7}$$

so that the Hermitian form (A7) is positive definite and

$$\det (h_{kj} + \mu\delta_{kj}) \neq 0$$

Then the system of equations (A6) does have a solution $a[\mu]$.

From the complex conjugates of (A6) and (A7) we find that

$$(a[\mu], e) \neq 0 \tag{A8}$$

so that $a[\mu]$ is not orthogonal to e. Hence an arbitrary vector a may be written in the form

$$a = \alpha a[\mu] + b \tag{A9}$$

where α is a complex scalar and b is a vector orthogonal to e, so that

$$(b, e) = 0 \tag{A10}$$

Note that if α were zero, we should have

$$f_0(\mathbf{u}_0) = (b, e) = 0 \tag{A11}$$

corresponding to no radiation in the direction $\mathbf{u}_0$; so we may assume $\alpha \neq 0$. Since $G(a)$ and $K(a)$ are quotients of Hermitian forms in the components of a, their values do not change when a is multiplied by a constant. Hence it suffices to consider excitations of the form

$$a = a[\mu] + b \tag{A12}$$

where b satisfies (A10). Such excitations leave (a, e) constant, so we need only study the behavior of (a, a) and (Ha, a) as functions of b.

To satisfy the condition

$$K(a) \leqq K(a[\mu]) \tag{A13}$$

it is necessary that

$$(a[\mu], b) + (b, a[\mu]) + (b, b) \leqq 0 \tag{A14}$$

Then we have

$$\begin{aligned}
(Ha, a) &= (Ha[\mu], a[\mu]) + (Ha[\mu], b) + (Hb, a[\mu]) + (Hb, b) \\
&= (Ha[\mu], a[\mu]) + (Ha[\mu], b) + (b, Ha[\mu]) + (Hb, b) \\
&= (Ha[\mu], a[\mu]) + (e, b) - \mu(a[\mu], b) \\
&\qquad + (b, e) - \mu(b, a[\mu]) + (Hb, b) \\
&\geqq (Ha[\mu], a[\mu]) + \mu(b, b) + (Hb, b) \\
&\geqq (Ha[\mu], a[\mu])
\end{aligned} \tag{A15}$$

where the second step follows because H is Hermitian, the third step on account of (A6), and the fourth step on account of (A10) and (A14). The equality signs in the last two steps hold if and only if $b = 0$. Hence the only possible solution to the problem of maximizing $G(a)$ under the condition $K(a) \leqq K(a[\mu])$ is given by $a = \alpha a[\mu]$, where α is a complex scalar. This proves Theorem I.

Theorem II. Since H is a positive definite Hermitian matrix, it has n positive real eigenvalues $\lambda_1, \lambda_2, \cdots, \lambda_n$, and n linearly independent eigenvectors $v^{(1)}, v^{(2)}, \cdots, v^{(n)}$, which satisfy

$$Hv^{(r)} = \lambda_r v^{(r)}, \qquad r = 1, 2, \cdots, n \tag{A16}$$

The eigenvectors may be taken as orthonormal, so that

$$(v^{(r)}, v^{(s)}) = \delta_{rs} \tag{A17}$$

We may expand the right-hand side of equations (A6) in terms of the vectors $v^{(r)}$; thus

$$e = \sum_{r=1}^{n} E_r v^{(r)} \tag{A18}$$

where

$$E_r = (e, v^{(r)}), \qquad r = 1, 2, \cdots, n \tag{A19}$$

Assuming a similar expansion of the solution of (A6), write

$$a[\mu] = \sum_{r=1}^{n} c_r v^{(r)} \tag{A20}$$

Substituting (A18) and (A20) into (A6) and using (A16) and (A17), we find without difficulty,

$$c_r = E_r/(\lambda_r + \mu) \tag{A21}$$

We can now write down expressions for the Hermitian forms occurring in $G(a)$ and $K(a)$. Using (A18), (A20), and (A21), and the orthogonality condition (A17), we obtain

$$(a[\mu], e) = \sum_{r=1}^{n} |E_r|^2/(\lambda_r + \mu) \tag{A22}$$

$$(a[\mu], a[\mu]) = \sum_{r=1}^{n} |E_r|^2/(\lambda_r + \mu)^2 \tag{A23}$$

$$(Ha[\mu], a[\mu]) = \sum_{r=1}^{n} \lambda_r |E_r|^2/(\lambda_r + \mu)^2 \tag{A24}$$

Note that all three expressions are positive real-valued, and that

$$d(a[\mu], e)/d\mu = -(a[\mu], a[\mu]) \tag{A25}$$

From the preceding expressions and the definition (A5) of $K(a)$, we have

$$\frac{dK(a[\mu])}{d\mu} = \frac{1}{(a[\mu], e)^3}\left[(a[\mu], e)\frac{d(a[\mu], a[\mu])}{d\mu} + 2(a[\mu], a[\mu])^2\right] \tag{A26}$$

The first factor on the right is positive, and the terms in square brackets become

$$-2\sum_{r=1}^{n}\frac{|E_r|^2}{(\lambda_r+\mu)}\sum_{s=1}^{n}\frac{|E_s|^2}{(\lambda_s+\mu)^3} + 2\left[\sum_{r=1}^{n}\frac{|E_r|^2}{(\lambda_r+\mu)^2}\right]^2 \leqq 0 \tag{A27}$$

by Schwarz's inequality. Hence $K(a[\mu])$ is a nonincreasing function of μ.

Now let $\mu_2 > \mu_1$ and let K_1, K_2, G_1, G_2 denote $K(a[\mu_1])$, $K(a[\mu_2])$, $G(a[\mu_1])$, $G(a[\mu_2])$. By Theorem I

$$G_2 = \max_{K(a)\leqq K_2} G(a) \leqq \max_{K(a)\leqq K_1} G(a) = G_1 \tag{A28}$$

since we have just proved $K_2 \leqq K_1$. This completes the proof of Theorem II.

Theorem III. To find the minimum possible value of $K(a)$, let

$$a = e + b \tag{A29}$$

where

$$(b, e) = 0 \tag{A30}$$

The coefficient of e may be taken as unity, since if it were zero, $K(a)$ would be infinite. When a is given by (A29), (a, e) has the constant value

$$(a, e) = n \tag{A31}$$

and making use of (A30),

$$(a, a) = n + (b, b) \geqq n \tag{A32}$$

Referring back to (A5), we see that $K(a)$ achieves the minimum value $1/n$ when $a = e$; this is Theorem III.

Theorem IV. To maximize $G(a)$ when there is no restriction on $K(a)$, set $\mu = 0$ in the system of equations (A6). We do not require (A13) or (A14), but from (A15) we still get

$$\begin{aligned}(Ha, a) &= (Ha[0], a[0]) + (e, b) + (b, e) + (Hb, b)\\ &= (Ha[0], a[0]) + (Hb, b)\\ &\geqq (Ha[0], a[0])\end{aligned} \tag{A33}$$

which shows that $G(a)$ is maximum when $a = a[0]$, and thus proves Theorem IV.

Theorem V. Setting $\mu = 0$ in (A23) and (A24) and using (A19), we obtain from (A4) an expression for the maximum gain in any direction $\mathbf{u}$, namely

$$\Gamma(\mathbf{u}) = \sum_{r=1}^{n} \lambda_r^{-1} \mid E_r \mid^2 = \sum_{r=1}^{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \lambda_r^{-1} e_s v_s^{(r)*} e_t^* v_t^{(r)} \tag{A34}$$

The dependence on $\mathbf{u}$ appears in the components of e, where

$$e_k = \exp(-i\beta \mathbf{r}_k \cdot \mathbf{u}), \qquad k = 1, 2, \cdots, n \tag{A35}$$

Averaging over all directions in space gives

$$\frac{1}{4\pi} \int_S e_s e_t^* \, d\Omega = \frac{\sin \beta r_{st}}{\beta r_{st}} = h_{st} \tag{A36}$$

the identification with the matrix element h_{st} being made from equation (29) on the assumption that the elements of the array are isotropic and the conventional definition of gain is used. It follows now from (A16) and (A17) that

$$\begin{aligned} \frac{1}{4\pi} \int_S \Gamma(\mathbf{u}) \, d\Omega &= \sum_{r=1}^{n} \sum_{s=1}^{n} \sum_{t=1}^{n} \lambda_r^{-1} v_s^{(r)*} h_{st} v_t^{(r)} \\ &= \sum_{r=1}^{n} \sum_{s=1}^{n} \lambda_r^{-1} \delta_{rs} \lambda_s = n \end{aligned} \tag{A37}$$

which proves Theorem V.

Bell System Technical Papers Not Published in this Journal

ANDERSON, R. E. D.[1]

A Magnetically Regulated Portable Battery Charger, A.I.E.E. Commun. and Electronics, **16,** pp. 607–610, Jan., 1955.

BOWN, R.[1]

Transistor as an Industrial Research Episode, Sci. Monthly, **80,** pp. 40–46, Jan., 1955.

BOYLE, W. S.,[1] and KISLIUK, P.[1]

Departure From Paschen's Law of Breakdown in Gases, Phys. Rev., **97,** pp. 255–259, Jan. 15, 1955.

BRATTAIN, W. H.,[1] and GARRETT, C. G. B.[1]

Surface Properties of Semiconductors, Physica, **20,** pp. 885–892, Nov., 1954.

BREIDT, P., JR., see GREINER, E. S.

BURTON, J. A.[1]

Impurity Centers in Ge and Si, Physica, **20,** pp. 845–854, Nov., 1954.

CHAPANIS, A., see SCALES, E. M.

CUTLER, C. C.[1]

The Regenerative Pulse Generator, Proc. I.R.E., **43,** pp. 140–149, Feb., 1955.

DECOSTE, J. B.,[1] and WALLDER, V. T.[1]

Weathering of Polyvinyl Chloride Wire and Cable Applications, Ind. and Eng. Chem., **47,** pp. 314–322, Feb., 1955.

[1] Bell Telephone Laboratories, Inc.

DEWALD, J. F.[1]

Theory of the Kinetics of Formation of Anode Films at High Fields, Electrochem. Soc. J., **102,** pp. 1–6, Jan., 1955.

FINE, M. E.[1]

Apparatus for Measuring the Elastic Moduli and Internal Friction of Solids from 1.7 to above 77°K and Some Values for α-Quartz, Rev. Sci. Instr., **25,** pp. 1188–1190, Dec., 1954.

FINE, M. E.,[1] and KENNEY, N. T.[1]

Low-Temperature Acoustic Relaxation in Ni-Fe Ferrites, Phys. Rev., **96,** pp. 1487–1488, Dec. 15, 1954.

FRANKE, H. C.[1]

Noise Measurements on Telephone Circuits, Tele-Tech **14,** pp. 85–87, 152–153, Mar., 1955.

FROEHLICH, FRITZ E.,[1] and SITTE, KURT[4]

Mean Free Path for Shower Production by High-Energy Pi Mesons, Phys. Rev., **97,** pp. 151–159, Jan. 1, 1955.

FROEHLICH, FRITZ E., see SITTE, KURT

FRYBURG, G. C., see TUMBORE, T.

FULLER, C. S., see REISS, HOWARD

GARRETT, C. G. B., see BRATTAIN, W. H.

GELLER, S.[1]

The Rhodium-Germanium System I: The Crystal Structures of Rh_2Ge, Rh_5Ge_3 and RhGe, Acta Crystallagraphica, **8,** pp. 15–21, Jan. 10, 1955.

GREINER, E. S.,[1] and BREIDT, P., JR.[1]

Melting Point of Germanium and the Constitution of Some Ge-Ga Alloys, J. Metals, **7,** pp. 187–188, Jan., 1955.

GREINER, E. S.[1]

The Plastic Deformation of Germanium and Silicon by Torsion, J. Metals, **7,** pp. 203–205, Jan., 1955.

[1] Bell Telephone Laboratories, Inc.
[4] Syracuse University, N. Y.

GOSS, A. J., see LOGAN, R. A.

HAMILTON, B. H.[1]

Some Applications of Semiconductor Devices in the Feedback Loop of Regulated Metallic Rectifiers, A.I.E.E. Commun. and Electronics, **16,** pp. 640–645, Jan., 1955.

HAMILTON, B. H.[1]

Semiconductor Devices in Regulated Rectifiers, Elec. Eng., **74,** p. 149, Feb., 1955.

HARTLEY, R. V. L.[1]

A New System of Logarithmic Units, Elec. Eng., **74,** pp. 135–137, Feb., 1955.

HAYNES, J. R., see HORNBECK, J. A.

HERRING, C.[1]

Theory of the Thermoelectric Power of Semiconductors, Phys. Rev., **96,** pp. 1163–1187, Dec. 1, 1954.

HERRING, C., see PEARSON, G. L.

KARP, A.[1]

Traveling-Wave Tube Experiments at Millimeter Wavelengths with a New, Easily Built, Space-Harmonic Circuit, Proc. I.R.E., **43,** pp. 41–46, Jan., 1955.

KELLY, M.J.[1]

Role of Industrial Research and Development in Society, Ind. Labs., **5,** pp. 6–10, Dec., 1954.

KENNEY, N. T., see FINE, M. E.

HORNBECK, J. A.,[1] and HAYNES, J. R.[1]

Trapping of Minority Carriers in Silicon. Part I: P-Type Silicon, Phys. Rev., **97,** pp. 311–321, Jan. 15, 1955.

HROSTOWSKI, H. J.,[1] and TANENBAUM, M.[1]

Recent Work On Group III Antimonides and Arsenides, Physica, **20,** pp. 1065–1067, Nov., 1954.

[1] Bell Telephone Laboratories, Inc.

KISLIUK, P., see BOYLE, W. S.

LAW, J. T.[1]

The Adsorption of Water Vapor on Ge and GeO_2, J. Phys. Che mistry **59,** p. 76, Jan., 1955.

LOGAN, R. A.,[1] GOSS, A. J.,[1] and SCHWARTZ, M.[1]

Semiconductor Devices made with Single Crystal Germanium Silicon Alloys, Letter to the Editor, J. Appl. Phys., **25,** 1551–1552, Dec., 1954.

MASON, W. P.[1]

Aging of the Properties of Barium Titanate and Related Ferroelectric Ceramics, J. Acous. Soc. Am., **27,** pp. 73–85, Jan., 1955.

MASON, W. P.[1]

Ultrasonic Attenuation In Normal Conducting Lead at Low Temperatures, Letter to the Editor, Phys. Rev., **97,** p. 557, Jan. 15, 1955.

MATTHIAS, B. T.[1]

Empirical Relation Between Superconductivity and the Number of Valence Electrons per Atom, Phys. Rev., **97,** pp. 74–76, Jan. 1, 1955.

MERRILL, J. L., JR.,[1] SMETHURST, J. O.,[1] and ROSE, A. F.[2]

Repeater Amplifies in Either Line Direction, Electronics, **28,** pp. 165–167, Jan., 1955.

MOORE, GEORGE, see WOOTEN, L. A.

NADELHAFT, IRVING, see SITTE, KURT

PEARSON, G. L.,[1] and HERRING, C.[1]

Magneto-resistance Effect and the Band Structure of Single Crysta Silicon, Physica, **20,** pp. 975–978, Nov., 1954.

PIERCE, J. R.[1]

The General Sources of Noise in Vacuum Tubes, I.R.E. Trans. P.G.E.D. ED-1, pp. 137–167, Dec. 1954.

RAISBECK, G.[1]

Definition of Passive Linear Networks in Terms of Time and Energy, J. Appl. Phys., **25,** pp. 1510–1515, Dec. 1954.

[1] Bell Telephone Laboratories, Inc.
[2] American Telephone and Telegraph Company

REISS, HOWARD,[1] and FULLER, C. S.[1]

Ionization and Solubility in Semiconductors, Letter to the Editor, Phys. Rev., **97,** pp. 559–560, Jan. 15, 1955.

RIORDAN, J.[1]

Discordant Permutations, Scripta Mathematica, **20,** pp. 14–23, March–June, 1954.

ROMNES, H. I.[2]

Requirements for Engineering in the Communications Operating Field, J. Eng. Ed., **45,** pp. 370–373, Dec., 1954.

RUEHLE, A. E., see WOOTEN, L. A.

SCALES, E. M.,[1] and CHAPANIS, A.[1]

Effect on Performance of Tilting the Toll-Operator's Keyset, J. Appl. Phys., **38,** pp. 452–456, Dec., 1954.

SCHWARTZ, M., see LOGAN, R. A.

SITTE, KURT,[4] FROEHLICH, FRITZ, E.,[1] and NADELHAFT, IRVING[4]

Electron Production in High-Energy Nuclear Interactions, Phys. Rev., **97:** pp. 166–172, Jan. 1, 1955.

SITTE, KURT, see FROEHLICH, FRITZ, E.

SUHL, H.[1]

Ferromagnetic Resonance in Nickel Ferrite Between One and Two Kilomegacycles, Letter to the Editor, Phys. Rev., **97,** pp. 555–557, Jan. 15, 1955.

TANENBAUM, M., see HROSTOWSKI, H. J.

TRUMBORE, F.,[1] and FRYBURG, G. C.[3]

Discussion of "The Electrolytic Preparation of Molybdenum from Fused Salts. III. Studies of Electrode Potentials", S. Senderoff and A. Brenner, J. Electrochemical Soc., **101,** p. 633, Dec., 1954.

---

[1] Bell Telephone Laboratories, Inc.
[2] American Telephone and Telegraph Company
[3] National Advisory Committee for Aeronautics
[4] Syracuse University, N. Y.

VONOHLSEN, L. H.[1]

The Small Signal Performance of the 416B Planar Triode Between 60–4000 Megacycles, I.R.E. Trans., P.G.E.D., p. 189, Dec., 1954.

WALLDER, V. T., see DECOSTE, J. B.

WOOTEN, L. A.,[1] RUEHLE, A. E.,[1] and MOORE, GEORGE, E.[1]

Evaporation of Barium and Strontium from Oxide-Coated Cathode, J. Appl. Phys., **26,** pp. 44–51, Jan.. 1955.

Recent Monographs of Bell System Technical Papers Not Published in This Journal*

ANDERSON, O. L., and STUART, D. A.

Activation Energy of Ionic Conduction in Silica Glasses, Monograph 2343.

BENNETT, W. R.

Sources and Properties of Electrical Noises, Monograph 2324.

BOZORTH, R. M.

Magnetostriction and Anistropy of Cobalt Crystals, Monograph 2314.

CRUSER, V. I.

Features of the AN/TRC-24 Radio Set, Monograph 2331.

EBERS, J. J., and MOLL, J. L.

Large-Signal Behavior of Junction Transistors, Monograph 2354.

FISK, J. B.

Acoustics in Communication, Monograph 2310.

HERRING, C.

Theory of the Thermoelectric Power of Semiconductors, Monograph 2387.

[1] Bell Telephone Laboratories, Inc.

* Copies of these monographs may be obtained on request to the Publication Department, Bell Telephone Laboratories, Inc., 463 West Street, New York 14, N. Y. The numbers of the monographs should be given in all requests.

VONOHLSEN, L. H.[1]

The Small Signal Performance of the 416B Planar Triode Between 60–4000 Megacycles, I.R.E. Trans., P.G.E.D., p. 189, Dec., 1954.

WALLDER, V. T., see DECOSTE, J. B.

WOOTEN, L. A.,[1] RUEHLE, A. E.,[1] and MOORE, GEORGE, E.[1]

Evaporation of Barium and Strontium from Oxide-Coated Cathode, J. Appl. Phys., **26,** pp. 44–51, Jan.. 1955.

Recent Monographs of Bell System Technical Papers Not Published in This Journal*

ANDERSON, O. L., and STUART, D. A.

Activation Energy of Ionic Conduction in Silica Glasses, Monograph 2343.

BENNETT, W. R.

Sources and Properties of Electrical Noises, Monograph 2324.

BOZORTH, R. M.

Magnetostriction and Anistropy of Cobalt Crystals, Monograph 2314.

CRUSER, V. I.

Features of the AN/TRC-24 Radio Set, Monograph 2331.

EBERS, J. J., and MOLL, J. L.

Large-Signal Behavior of Junction Transistors, Monograph 2354.

FISK, J. B.

Acoustics in Communication, Monograph 2310.

HERRING, C.

Theory of the Thermoelectric Power of Semiconductors, Monograph 2387.

[1] Bell Telephone Laboratories, Inc.

* Copies of these monographs may be obtained on request to the Publication Department, Bell Telephone Laboratories, Inc., 463 West Street, New York 14, N. Y. The numbers of the monographs should be given in all requests.

HOFFMANN, J. P., HUBER, G. H., MILLER, W. F., and SCHRAMM, C. W.
New Military Carrier Telephone Systems, Monograph 2350.

HUBER, G. H., see HOFFMANN, J. P.

JENSEN, A. G.
The Evolution of Modern Television, Monograph 2363.

MCLEAN, D. A., and WEHE, H. G.
Minature Lacquer Film Capacitors, Monograph 2384.

MILLER, W. F., see HOFFMANN, J. P.

MOLL, J. L.
Large-Signal Transient Response of Junction Transistors, Monograph 2383.

MOLL, J. L., see EBERS, J. J.

NORDAHL, J. G.
A New Military Radio Relay System, Monograph 2330.

PIERCE, J. R.
Some Recent Advances in Microwave Tubes, Monograph 2353.

RAISBECK, G.
Passive Linear Networks in Terms of Time and Energy, Monograph 2352.

REISS, H.
Mathematical Methods for Zone-Melting Processes, Monograph 2297.

RICE, S. O.
Distribution of a Sum of *n* Sine Waves, Monograph 2365.

SCHRAMM, C. W., see HOFFMANN, J. P.

STUART, D. A., see ANDERSON, O. L.

TIEN, P. K.
Focusing of a Long Cylindrical Electron Stream, Monograph 2293.

WEHE, H. G., see MCLEAN, D. A.

Contributors to this Issue

WILLIAM R. BENNETT, B.S. in E.E., Oregon State College, 1925; M.A., Columbia University, 1928; Ph.D., Columbia, 1949. Bell Telephone Laboratories, 1925–. His early Laboratories projects included work on wire transmission problems, particularly the development of terminal apparatus in the voice and telegraph range, the design of circuits for television, and submarine cable telephony. Concerned with the coaxial cable in 1935, he spent several years working on the requirements and measuring techniques applicable to the load rating of multi-channel repeaters. His work during World War II was directed to a number of military projects. Since then he has concentrated on pulse code modulation and general transmission problems. Member of the A.I.E.E., I.R.E., The American Physical Society, Tau Beta Pi, Eta Kappa Nu and Sigma Xi.

PAUL W. BLYE, S.B. in E.E., Massachusetts Institute of Technology, 1919, American Telephone and Telegraph Company, 1919–34; Bell Telephone Laboratories, 1934–. His early work at the A.T. and T. Co. and the Laboratories concerned the development of special testing apparatus and methods for inductive coordination of power lines and telephone circuits, and later, studies on interference prevention. During World War II, he supervised a number of military projects, including the laying of field wire from airplanes. In 1946 he was named Transmission Systems Engineer, and in 1951 to his present post as Director of Transmission Engineering at the Laboratories. Served on the joint Sub-Committee on Development and Research of the Edison Electric Institute and as a consultant to N.D.R.C. from 1941–45. Fellow of the A.I.E.E., and Senior Member of the Institute of Radio Engineers.

G. M. BOUTON, Ch.E., Brooklyn Polytechnic Institute, 1926; Bell Telephone Laboratories, 1926–. Over the years he has been engaged primarily in research and development studies on metals and alloys. He is an authority on cable sheath alloys and has published several technical articles relating to them. He has been granted numerous patents on cable sheath alloys and solders. He is a member of the American Society for Metals and the American Institute of Mining and Metallurgical Engineers.

OLIVER H. COOLIDGE, A.B., Harvard College, 1922; New York Telephone Company, 1921–27; American Telephone and Telegraph Company, 1927–34; Bell Telephone Laboratories, 1934–. After six years in plant maintenance methods work with the New York Company, he joined the Development and Research Department of the A. T. and T. Co. in 1927; this department was transferred to Bell Laboratories in 1934. For the past twenty-eight years his work has been both experimental and theoretical in the field of transmission engineering. He was especially concerned with problems of low-energy interference prevention, such as noise and crosstalk in voice, carrier, radio and video systems. During World War II he served as a radar maintenance instructor in the Laboratories' School for War Training. Since 1949 his work has been largely on problems of quality and standards of local transmission.

HAROLD E. CURTIS, B. S. and M.S., Massachusetts Institute of Technology, 1929. He joined the Department of Development and Research of the American Telephone and Telegraph Company in 1929, and was transferred to the Bell Telephone Laboratories in 1934. Mr. Curtis has been concerned with transmission problems related to multi-channel carrier telephony and television. He has been involved particularly in studies of transmission engineering aspects of microwave radio relay systems. His work at the Laboratories has also included pioneering transmission studies of the coaxial cable, the shielded pair and quad, and the waveguide. Mr. Curtis holds ten patents relating to carrier telephony.

EDGAR NELSON GILBERT, B.S., Queens College, 1943; Ph.D., Massachusetts Institute of Technology, 1948; M. I. T. Radiation Laboratory, 1944–46. Mr. Gilbert joined Bell Telephone Laboratories in 1948 and was concerned at first with studies of information theory, and later with switching theory. He is presently a member of a group concerned with probability and information theory, and with discrete systems. Mr. Gilbert is a member of the American Mathematical Society.

HAYDEN W. EVANS, B.A., Ohio Wesleyan University, 1934; B.S. in E.E., University of Michigan, 1936; Bell Telephone Laboratories, 1936–. During his early association with the Laboratories, Mr. Evans was concerned with transmission engineering problems on open wire and cable circuits, including transmission of television on open wire lines, open-wire crosstalk, cable crosstalk, and coaxial cable transmission. Later, he was engaged in the development of radar, radar test equipment, and countermeasures equipment. Following the war, he was concerned with the pri-

mary engineering of broad-band radio relay systems, and radio transmission engineering, including mobile radio problems and radio relay systems. He is a senior member of the I.R.E., a member of the Acoustical Society of America, Tau Beta Pi, Sigma Xi and several other honorary fraternities.

JOHN HERBERT HEISS, B.S. in Chemical Engineering, Newark College of Engineering, 1942. Mr. Heiss joined Bell Telephone Laboratories in 1934 and was concerned at first with experimental wire-coating procedures and test methods. Later he was involved in experimental production of high polymers and examination of their physical properties, and also with research studies on solution properties of high polymers. For some time, Mr. Heiss worked on studies of polymer rheology. More recently, he has been engaged in polymer mechanics studies in the Chemical Research Department. He is a member of the American Chemical Society.

ARTHUR W. HORTON, JR., A.B. Princeton, 1920; E.E., Princeton, 1922. Mr Horton spent the summer of 1921 as an engineering student in the Physical Laboratories of the Western Electric Company, and on his graduation from college, joined the Transmission Department of that company. He became a member of the Research Department of Bell Telephone Laboratories on its organization in 1925. He has been concerned with telegraph, picture and television transmission and voice-operated equipment, such as the transatlantic radio terminal and echo suppressors. During World War II he worked on an underwater sound detection system for submarine mines and on the development of the Mark 8 computer for antiaircraft fire control. He was a member of the National Defense Research Committee and the recipient of a Certificate of Exceptional Service to Naval Ordnance Development. From 1945 to 1946 he worked on indicators for naval fire control radar. He is presently in charge of groups working on military systems switching research. Senior Member of the Institute of Radio Engineers, member of the American Physical Society, American Institute of Physics, Association for Computing Machinery.

H. R. HUNTLEY, B.S. in E.E., University of Wisconsin, 1921; Wisconsin Telephone Company, 1917–1930, except for a leave of absence to complete education begun earlier at Leland Stanford University and continued at the University of Wisconsin. Leaving the Wisconsin Telephone Company where he was Transmission Engineer, Mr. Huntley came

to the Foreign Wire Relations Section of the Operating and Engineering Department of American Telephone and Telegraph Company in 1930. In 1942 he transferred to the Transmission Section and has been Transmission Engineer since 1951.

Samuel P. Morgan, B.S., California Institute of Technology, 1943; M.S., California Institute of Technology, 1944; Ph.D., California Institute of Technology, 1947; Bell Telephone Laboratories, 1947–. A research mathematician, Dr. Morgan specializes in electromagnetic theory. He has been particularly concerned with problems of waveguide and coaxial cable transmission and microwave antenna theory. Member of the American Physical Society, the Institute of Radio Engineers, Tau Beta Pi, and Sigma Xi.

George, S. Phipps, B.S. in Electrochemical Engineering, Pennsylvania State College, 1930; M.S. in Metallurgy, Columbia University, 1939. Mr. Phipps joined the technical staff of Bell Telephone Laboratories in 1930 on graduation from college. Throughout his telephone career he was chiefly engaged in metallurgical research on low melting alloys, solders and related material. He was responsible for the selection or development of many of the solders now used in the Bell System. He was a member of the American Society for Metals. Mr. Phipps died February 16, 1955, after a brief illness.

G. A. Pullis, Western Electric Company, 1920–1924; Bell Telephone Laboratories 1925–. Mr. Pullis joined the Laboratories in 1920 and his early work involved the testing and development of transmission instruments. Since 1928 he has been concerned with the design and development of toll signaling arrangements. He has also worked on the design of remote control, alarm and order wire arrangements for the TD-2 microwave radio relay system. He is presently in charge of a group working on toll signaling and television program switching.

Eugene D. Reed, B.Sc., University of London, 1941; M.S. and Ph.D. in E.E., Columbia University, 1947 and 1953. U. S. Army, 1944–46; Bell Telephone Laboratories, 1947–. Dr. Reed is engaged in the development and design of microwave oscillators. Member of the Institute of Radio Engineers and Sigma Xi.

Stephen O. Rice, B.S., Oregon State College, 1929; California Institute of Technology, Graduate Studies, 1929–30 and 1934–35; Bell Tele-

phone Laboratories, 1930–. In his first years at the Laboratories, Mr. Rice was concerned with non-linear circuit theory, with special emphasis on methods of computing modulation products. Since 1935 he has served as a consultant on mathematical problems and in investigations of telephone transmission theory, including noise theory, and applications of electromagnetic theory. Fellow, I.R.E.

H. EARLE VAUGHAN, B.S. in C.E., Cooper Union, 1933. Bell Telephone Laboratories, 1928–. In his early years with the Laboratories, Mr. Vaughan worked on voice-operated devices. In 1937 he turned to studies of the effect of speech and noise on voice-frequency signaling systems and continued in this type of work until World War II when he concentrated on the development of anti-aircraft computers, fire-control radar and other military projects. Since the war he has been engaged in research on switching systems and high speed signaling devices. He is presently in charge of groups specializing in research on electronic switching systems. He is a Senior Member of the I.R.E.

IRWIN WELBER, B.S. in E.E., Union College, 1948; M.E.E., Rensselaer Polytechnic Institute, 1950; R.P.I. instructor in electrical engineering, 1948–1950; Bell Telephone Laboratories, 1950–. After he completed studies with the Laboratories' Communications Development Training Program, Mr. Welber was assigned to a group working on TD-2 automatic switching. He is currently concerned with TD-2 equalization. Associate Member I.R.E., A.I.E.E., Associate Member Sigma Xi.