THE BELL SYSTEM

Technical Journal

DEVOTED TO THE SCIENTIFIC AND ENGINEERING ASPECTS OF ELECTRICAL COMMUNICATION

VOLUME XXXIII NOVEMBER 1954 NUMBER 6

THE BELL SYSTEM TECHNICAL JOURNAL

THE BELL SYSTEM TECHNICAL JOURNAL is published six times a year by the American Telephone and Telegraph Company, 195 Broadway, New York 7, N. Y. Cleo F. Craig, President; S. Whitney Landon, Secretary; John J. Scanlon, Treasurer. Subscriptions are accepted at $3.00 per year. Single copies are 75 cents each. The foreign postage is 65 cents per year or 11 cents per copy. Printed in U. S. A.

THE BELL SYSTEM
TECHNICAL JOURNAL

VOLUME XXXIII NOVEMBER 1954 NUMBER 6

Waveguide as a Communication Medium

By S. E. MILLER

(Manuscript received March 23, 1954)

The circular electric wave in round metallic tubing has an attenuation coefficient which decreases as the frequency of operation is increased. A corollary to this behavior is the fact that any preselected attenuation coefficient can in theory be obtained in any predetermined diameter of pipe through the choice of a suitably high carrier frequency. The attenuation which is characteristic of microwave radio repeater links, about 2 db/mile, is in theory attainable in a copper pipe of about 2" diameter using a carrier frequency near 50,000 mc.

Scale-model transmission experiments, conducted at 9,000 mc, showed average transmission losses about 50 per cent above the theoretical value. These extra losses were due to (1) roughness of the copper surface and (2) transfer of power from the low-loss mode to other modes which can also propagate in the pipe.

The latter effect may have serious consequences on signal fidelity because power will transfer (at successive waveguide imperfections) from the signal mode to unused modes and, after a time delay, back to the signal mode. This effect has been studied experimentally and theoretically, and it is concluded that (1) either mode filters must be inserted periodically to absorb the power in the unused modes of propagation, or (2) the medium itself must be modified to continuously provide large attenuations for the unused modes of propagation. The latter approach is attractive in that it also provides a solution to the problem of bending this form of low-loss guide.

The general outlook, based on present knowledge, is that a waveguide system might transmit baseband widths as large as 100 to 500 mc using a rugged modulation method such as PCM. Some form of regeneration is likely to be required at each of the repeaters, which may be spaced on the order of 25 miles. A total rf bandwidth of about 40,000 mc may be available in a single guide.

TABLE OF CONTENTS

INTRODUCTION

The circular electric wave in round metallic tubing possesses a property so unique that some early research workers doubted the reality of the wave. This unique property is an attenuation coefficient which, in a given pipe, *decreases* without limit as the frequency of operation is *increased*. In parallel wire, coaxial, or ordinary waveguide lines the "skin effect" at the surface of the conductor causes the loss to increase as the frequency increases indefinitely, so the predicted circular-electric-wave loss characteristic aroused considerable interest as soon as it was discovered by S. A. Schelkunoff and G. C. Southworth in the early 1930's. Since that time considerable work has been done at Holmdel to explore the reality of the circular electric wave and to evaluate its usefulness to the Bell System. It is the purpose of this paper to report on the status of this work and to give a description of some of the basic characteristics of circular electric wave propagation.

The Bell System is interested in knowing whether waveguide can be used as a long distance communication medium in the manner in which coaxial cable or the radio relay system is now employed. Our interest in long distance waveguides is due in part to the fact that radio-wave propagation through the atmosphere becomes progressively more severely handicapped by rain, water vapor and oxygen absorptions at

frequencies above 12,000 mc. Use of the spectrum above the 10,000–20,000 mc region seems to require a sheltered transmission medium.

Circular electric wave transmission may also find application in short connecting links, such as between subscribers requiring very broad band circuits, between two central offices as a multi-channel carrier link, or between a radio relay antenna site and a somewhat remote transmitter-receiver location chosen for accessibility.

In each of these cases, the broad bands available in the microwave portion of the spectrum, the complete shielding afforded by waveguides generally, combined with the low-loss properties of the circular electric wave would seem to provide an ideal transmission medium. We therefore seek knowledge of the precision required in the waveguide and some indication of general system complexity to facilitate a judgment as to whether the cost will be competitive.

ORDINARY VERSUS CIRCULAR ELECTRIC WAVES

Let us approach a discussion of circular electric waves by considering their relation to the waveguides which are now used in our radio relay systems and which found widespread use in the radars of World War II. The vast majority of waveguides in commercial use now are rectangular in cross section and have dimensions large enough so that one and only one wave-type, usually called the "dominant mode", can propagate. To simplify this discussion, such waveguides will be called ordinary waveguides. Ordinary waveguides are analogous to coaxial or parallel-wire lines in many respects. Because only one mode can propagate, departures from an absolutely straight tube of constant cross section show up as reactance effects only. A dent in the side wall of the guide or of the coaxial, an abrupt change in cross section, or a twist or bend of the line all appear as non-dissipative reflection effects which may be cancelled at one frequency (or in one band of frequencies) by the addition of another compensating reactance at a point suitably located. A great many of the components used in ordinary waveguides, including the frequency selective filters, depend on such reactance cancellation effects in order to achieve satisfactory operation.

Since the techniques for employing ordinary waveguides have been thoroughly explored, it is natural to inquire as to whether we can use them for communication purposes. We do use ordinary waveguides in lengths of the order of 100 feet and more to connect the antennas and repeaters in the 4,000 mc (TD-2) radio relay system. The attenuation is excessive, however, for long-distance applications. The particular type

of brass rectangular waveguide used for TD-2 transmission lines has attenuation in excess of 50 db per mile, and use of the very best copper would only reduce the theoretical loss to about 40 db per mile. In order to reduce the loss in ordinary waveguides, just as in coaxial or parallel wire lines, one must go to lower frequencies. In particular, the theoretical loss at a carrier frequency near 1000 mc is about 2 db per mile, which is about the same as the transmitter-to-receiver attenuation in our radio relay systems. The waveguide in the 500–1000 mc region would have cross-sectional dimensions on the order of one foot, would be cumbersome to handle and would involve rather large material cost. In addition, it turns out that such a waveguide would be useful in signal bandwidths only a few mc wide as a result of delay distortion, which will be discussed further in the ensuing discussion. Thus, we have concluded that ordinary waveguide is not very attractive as a transmission medium over distances on the order of a mile or more.

It is true that the attenuation in any hollow metallic waveguide can be reduced to any desired extent at a given frequency by making the cross sectional area larger by a suitable factor. The penalty is that the transmission medium becomes capable of propagating energy in several characteristic ways, known as modes. The striking feature of a multimode transmission medium is that energy in one mode is entirely independent and unaltered by the presence or absence of energy in one of the other modes. This situation is sketched diagrammatically in Fig. 1. Energy can theoretically propagate between 1 and 1′, between 2 and 2′, and between 3 and 3′ at the *same time* and in the *same frequency band* without mutual interference. The separate modes represent independent transmission lines which occupy the same space. The distinguishing features of the various modes in a multi-mode waveguide are: (1) Velocity of propagation or phase constant, (2) Attenuation coefficient, and (3) Configuration of electric and magnetic field lines within the waveguide.

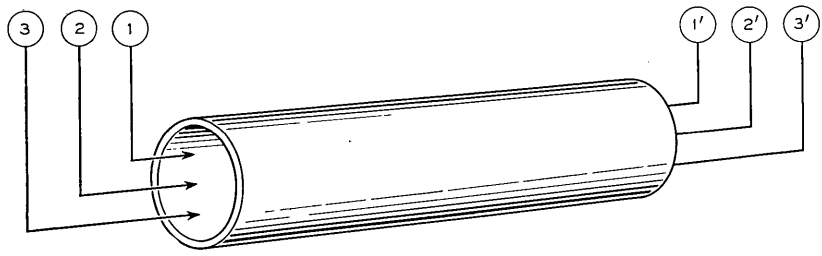

Fig. 1 — Diagram of multi-mode waveguide transmission.

The fact that it is necessary to use a waveguide whose dimensions are large enough to permit the existence of a number of modes has far-reaching influence on the research being discussed here. Practically, the independence between the various modes of propagation is limited by tolerances of various kinds. In the multimode waveguide, changes in cross section or bends or twists require design attention with regard to mode purity as well as with regard to impedance match, and it is not permissible to insert arbitrarily shaped probes or irises for impedance matching purposes as is the common practice in ordinary waveguides. This means that a complete new technique is required for the old components, such as frequency-selective filters, hybrids, and attenuators, as well as for a new series of components such as pure mode generators and mode filters.

THEORETICAL CHARACTERISTICS OF THE CIRCULAR ELECTRIC WAVE

Since it has been found necessary to use a waveguide in the multimode region in order to get the desired losses in a reasonable size waveguide, we may inquire as to which of the modes is best suited to our problem. At a given frequency the loss for any one of the modes may be reduced as much as is desired by making the cross sectional area of the guide large enough, but there is a mode for which the loss decreases with increasing guide size much more rapidly than for any other mode. This is the circular electric (TE_{01}) mode in straight round pipe. It turns out that no current flows in the direction of propagation in the metallic walls of a straight round pipe carrying the circular electric mode. It is the absence of current in the direction of propagation which permits the circular-electric-wave attenuation to decrease indefinitely as the frequency increases, and this difference between ordinary transmission lines and the circular-electric wave is further illustrated in Fig. 2. In the familiar parallel-wire line the electric field extends directly from one conductor to the other, resulting in charge accumulations at half-wave intervals along the axis of propagation and associated conduction currents in the copper wires. These conduction currents in the direction of propagation do not diminish as the frequency of operation increases, since they are associated with the energy transmitted to the end of the transmission line. With the circular electric wave the electric field lines close upon themselves, are always tangential to the conducting wall, and do not result in a charge accumulation on the walls due to the main energy flow. The wall currents which do flow are merely sufficient to prevent the propagating energy from spreading out as it would if the metallic walls

were not present, but these currents decrease rapidly with increasing frequency in a given waveguide size. Only in a straight circular pipe can all of the electric field lines close upon themselves, and only in the straight circular pipe does the attenuation approach zero at infinite frequency.*

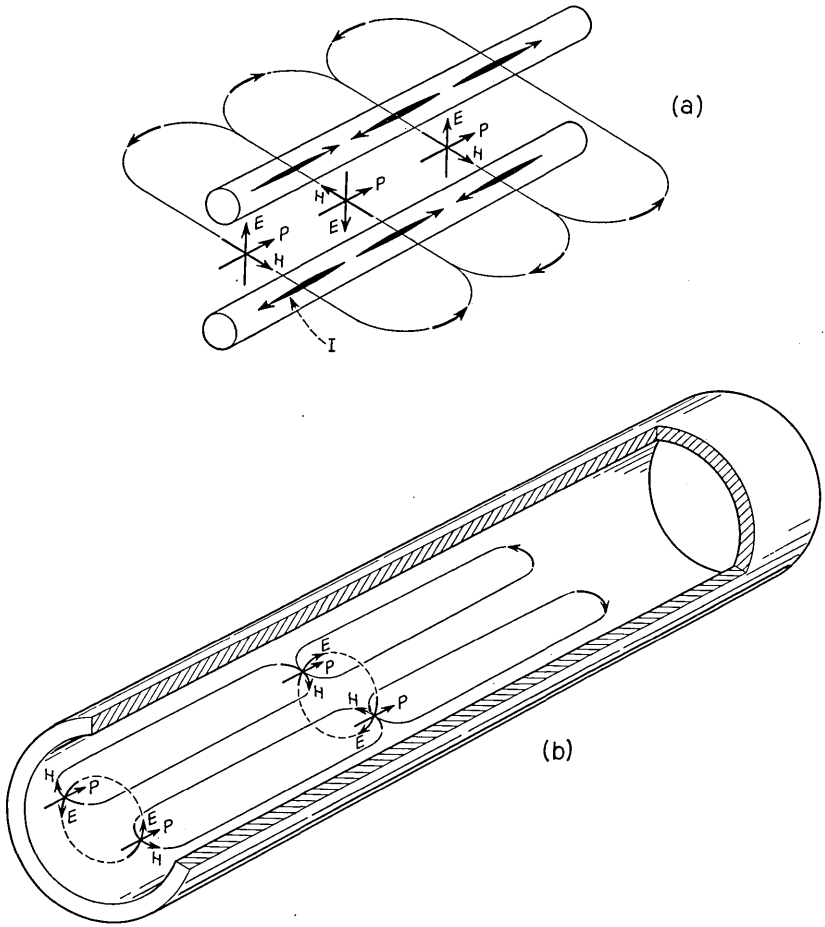

Fig. 2 — Sketch of the magnetic intensity (H), electric intensity (E) and Poynting Vector (P) for parallel-wire and circular-electric waveguides. Because the main energy flow (P) in the circular electric waveguide is associated with electric field lines that close on themselves and do not produce accumulations of charge on the metal walls, the wall currents and associated losses are very small.

* For further discussion, see G. C. Southworth, *Principles and Applications of Waveguide Transmission*, D. Van Nostrand, Inc. pp. 175–178.

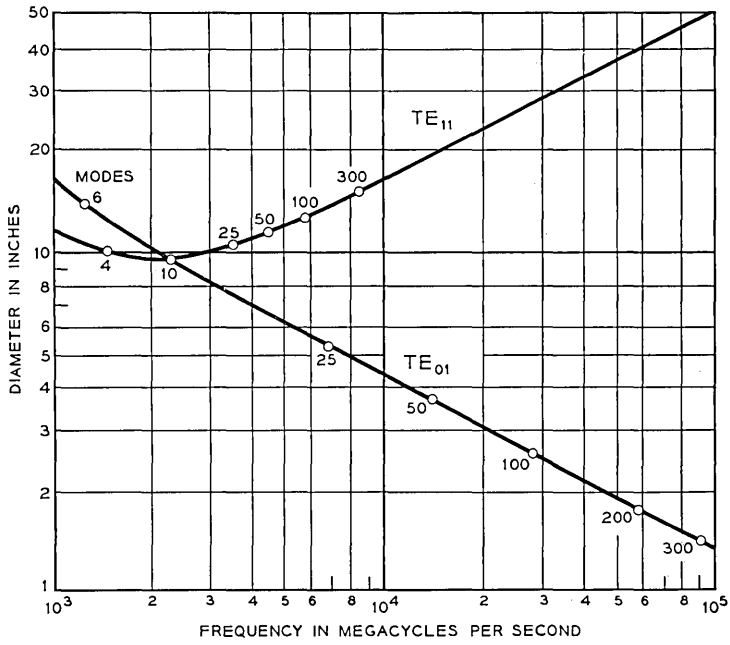

Fig. 3 — Round guide diameter versus frequency for attenuation of 2 db/mile.

As a consequence of the unusual loss versus frequency characteristic of the circular electric wave, the diameter required in order to achieve a given loss decreases as the carrier frequency increases. This is illustrated by the curves labeled TE_{01} in Figs. 3 and 4. All other waveguide modes (except higher-order circular-electric waves, TE_{0m}) have a characteristic of the general form sketched for the dominant wave (TE_{11}) also shown in Figs. 3 and 4. The longitudinal wall currents contribute a loss component which rises at increasing frequencies due to skin effect; this accounts for the positive slope of the TE_{11} curve at the right-hand side of Figs. 3 and 4. The negative slope of the TE_{01} curves and of the left-hand portion of the TE_{11} curves in Figs. 3 and 4 is a consequence of losses associated with the wall currents which prevent the wave from spreading as it would in an unbounded medium; these currents and the losses associated with them decrease as the operating frequency becomes farther removed from cut-off.

For a loss of 2 db per mile Fig. 3 shows that a waveguide 7″ in diameter is required in the frequency band near 4,000 mc where the TD-2 system operates. Whereas this may not be prohibitive in a connecting

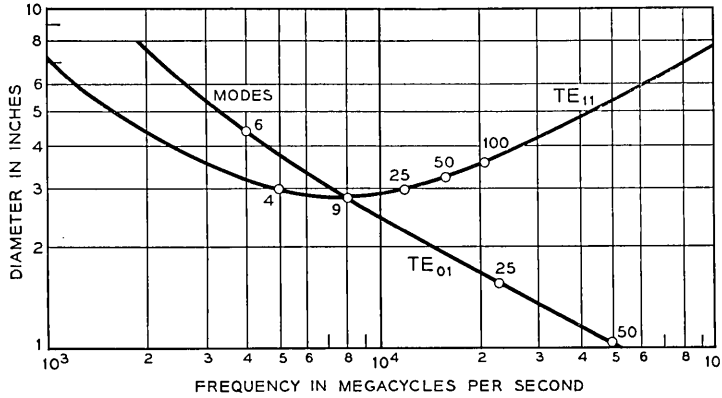

Fig. 4 — Round guide diameter versus frequency for attenuation of 13.2 db/mile (0.25 db/100 ft.).

link application, the waveguide size is definitely too large for long-distance application. In the vicinity of 50,000 mc, however, Fig. 3 shows that the required waveguide size is on the order of 2″, and this is comparable to the size of the present standard 8-pipe coaxial cable. From these simple calculations, it is evident that carrier frequencies in the vicinity of 50,000 mc or more are very desirable for long-distance waveguide applications in order to minimize the size of the waveguide.

Other reasons for wanting a high carrier frequency arise from a consideration of bandwidth. Any hollow conductor waveguide has a cutoff characteristic of the form sketched in Fig. 5. Above cutoff the group velocity approaches asymptotically to the velocity in an unbounded medium composed of the dielectric used in the waveguide. Because the group velocity varies across the frequency band, a signal being transmitted in a waveguide will experience delay distortion; the components transmitted at $f_0 \pm \Delta f$ (Fig. 5) would be delayed compared with their relation at the input to the line. When this delay is 180° a baseband signal of frequency Δf would be severely distorted regardless of the modulation method, and this condition may be regarded as an upper limit to the usable bandwidths in the waveguide unless correction for delay distortion is employed. This particular type of phase distortion has been analyzed in unpublished papers, first by D. H. Ring and later by S. Darlington. The work of Darlington leads to the following relation between the parameters of the waveguide and the baseband width f_B associated with the 180° phase difference noted above:

$$f_B = \frac{304 f^{1/2}(1 - v^2)^{3/4}}{v L^{1/2}} \text{ (cps)}$$

where f is the carrier frequency (cps), v is the ratio of the waveguide cutoff frequency to the carrier frequency, and L is the line length in miles.

From the above relation, the maximum baseband width available has been calculated for one mile of line as a function of carrier frequency with the loss held constant at 2 db per mile and 13.2 db per mile, and the results are plotted in Fig. 6. The conditions for these curves are directly comparable to those for which Figs. 3 and 4 were calculated. At a carrier frequency near 50,000 mc, the circular electric wave in 2-inch diameter pipe makes available a baseband width on the order of 500 mc for one mile of line, or 100 mc for 25 miles of line. At lower frequencies, with the waveguide enlarged to hold the loss constant, there is less bandwidth available.

The 13.2 db per mile condition (in a smaller diameter of waveguide) permits the use of approximately one-half the bandwidth available at the 2 db per mile condition.

The above design considerations are the basis for concluding that the most attractive communication possibility is the use of carrier frequencies on the order of 50,000 megacycles and associated waveguide

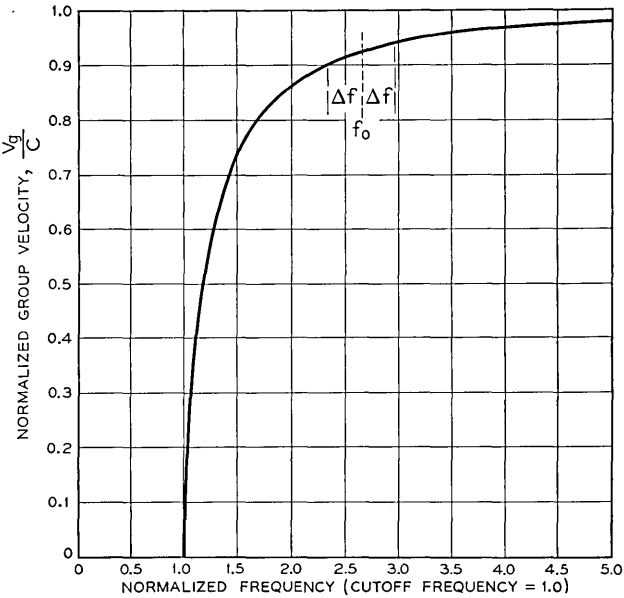

Fig. 5 — Normalized group velocity versus normalized frequency for hollow metallic waveguides.

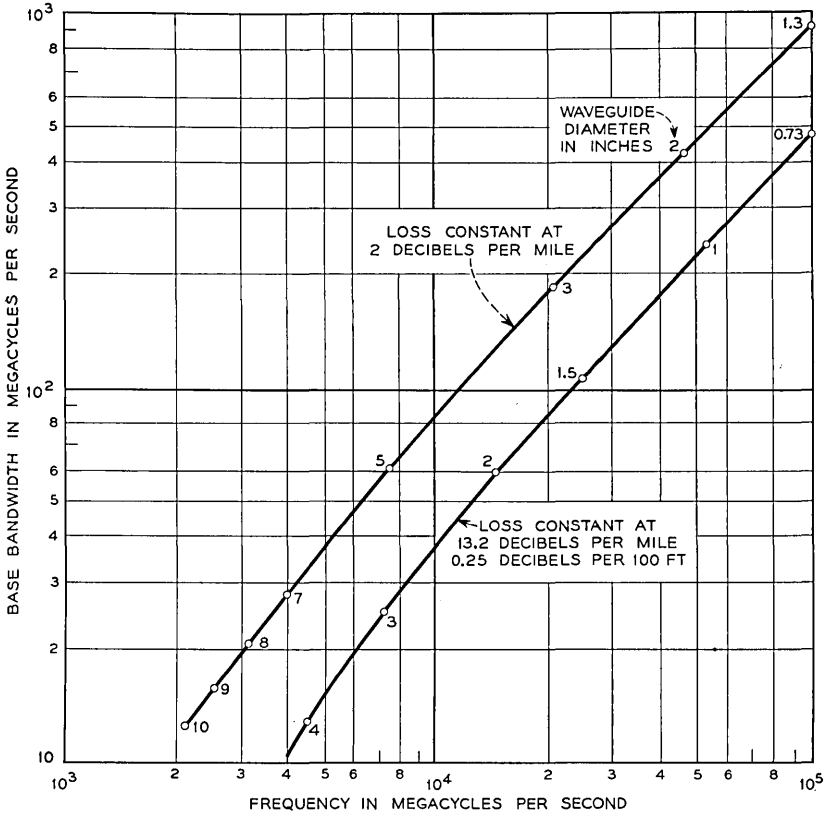

Fig. 6 — Base band width per channel versus frequency (for one-mile of waveguide) using the circular electric mode, loss held constant by varying pipe diameter.

diameters on the order of 1 to 2 inches. Having selected the general region which appears attractive, it is important to know the communication characteristics of a given waveguide. Figs. 7 and 8 show the attenuation and bandwidth characteristics respectively for a 2-inch round copper tube. The theoretical loss (Fig. 7) for the TE_{01} wave varies from 4 db per mile to 0.64 db per mile over the frequency range from 30,000 to 100,000 mc. The higher order circular electric waves have somewhat more attenuation but may be found useful as auxiliary transmission channels for short distances.

It is interesting and important, as will be pointed out later, to note that several of the higher order transverse electric waves have attenuation constants which may be within a factor of 3 to 6 times the attenuation for the circular electric wave. These waves, TE_{12} and TE_{13}, have

very little wall current flowing in the direction of propagation and therefore approximate the low-loss properties of the circular electric wave family. Under certain conditions of mode coupling, which will be described at a later point, it is undesirable for the medium to be able to propagate modes with attenuation coefficients comparable to that of the mode which is used for communication purposes.

SOME RESULTS OF TRANSMISSION EXPERIMENTS

Transmission experiments have been conducted on the 500-ft waveguide line shown in Fig. 9.* Supports for the line were set in concrete

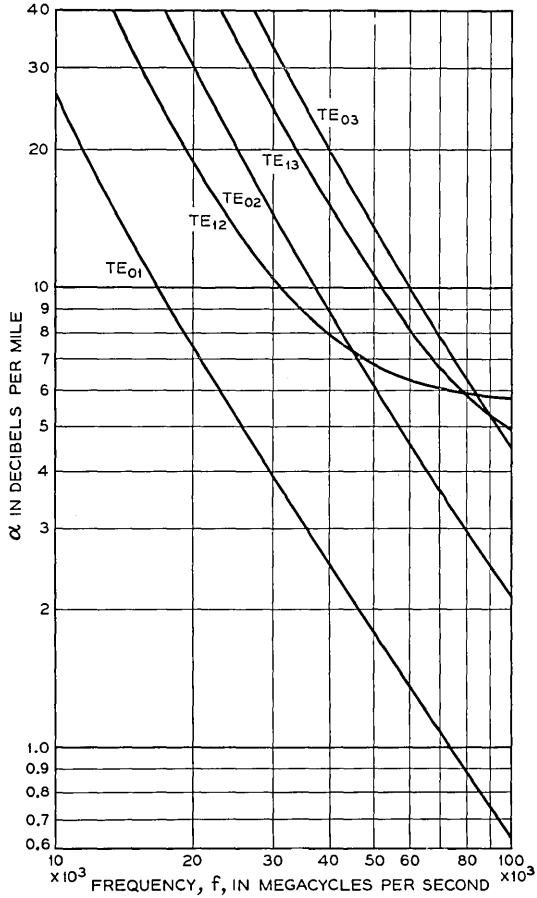

Fig. 7 — Attenuation versus frequency for a 2″ diameter round waveguide.

* This is the same line used for the work reported in Reference 1. Some of the experiments described in Reference 1 are also described here in order to furnish background for the new material.

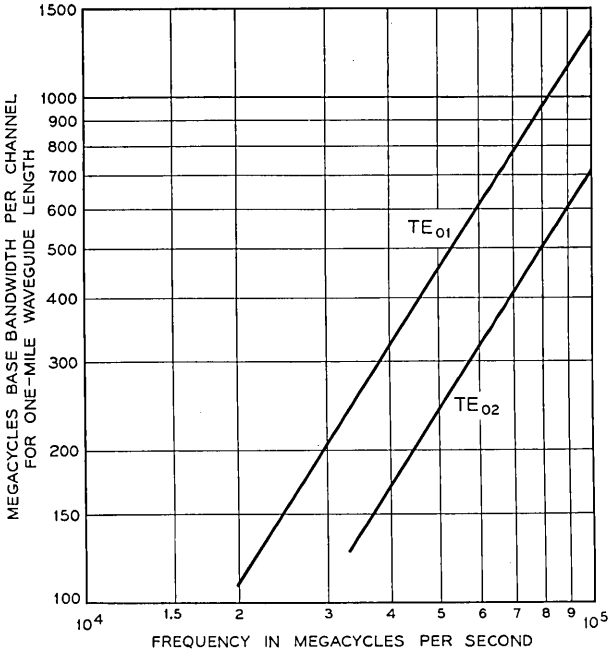

Fig. 8 — Base band width per channel versus frequency for a 2″ diameter pipe (one-mile waveguide length).

and optically aligned so as to provide a waveguide straight within about ⅛″ over its entire length. The philosophy behind this installation was the familiar one of providing for experimental purposes, as close to the ideal line as possible so that deviations could be created in a controlled manner. The inside diameter is about 4.73″, chosen to obtain the desired theoretical loss of about 2 db per mile at 9,000 mc, where measuring equipment was readily available. The difference between the major and minor inside diameters of the pipe was in the range 0.005″ to 0.008″ at the ends of the sections which averaged 20 ft in length. At the time this work was initiated, in 1946, generators of higher frequencies which would permit the use of smaller waveguides had not yet become available for use in this research.

Tests were conducted on this line using a technique due to A. C. Beck,[1] and involving the layout of equipment shown in Fig. 10. Short bursts of *RF* energy approximately 1/10 microsecond in duration, were injected into the line at intervals of about three hundred microseconds. Except for two small holes through which to couple to the transmitter and

receiver, the waveguide line was short-circuited at both ends. The injected $\frac{1}{10}$ microsecond pulse occupied at any instant a space interval of 100 feet and therefore this pulse, while travelling from one end to the other between the short circuits, produced at the receiver coupling hole spurts of energy corresponding to the time when the pulse passed the sending end. Each such pulse passing through the receiver coupling hole was amplified, detected, and placed as a vertical deflection on the oscilloscope. The horizontal deflection on the oscilloscope was a linear time

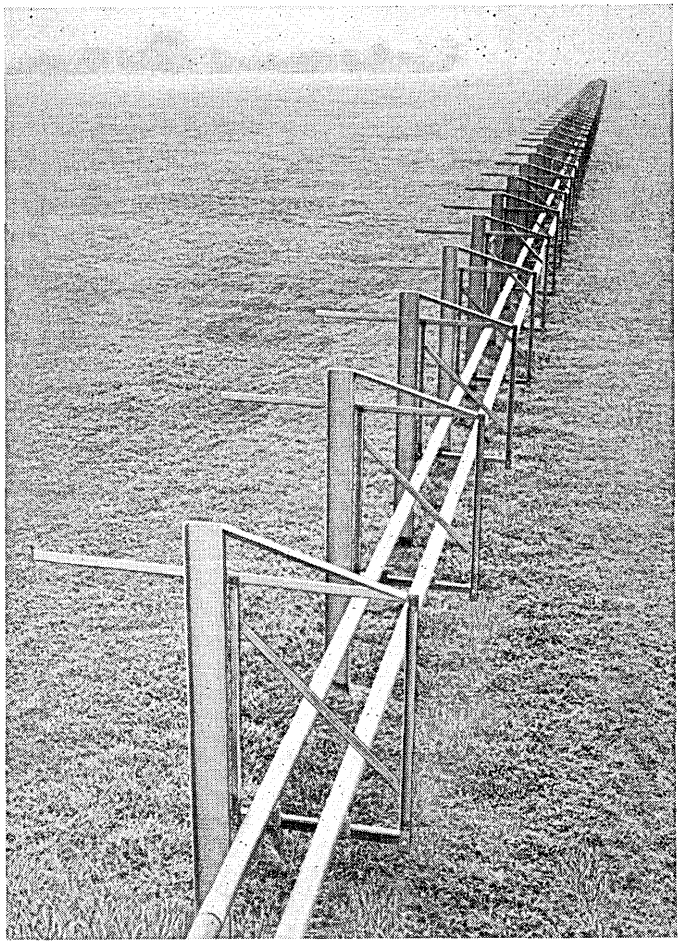

Fig. 9 — Experimental waveguide installation, 5″ diameter Holmdel line.

base having a duration of a few microseconds. In order to observe the received pulse after a selected number of trips back and forth down the waveguide, a variable delay was placed between the trigger and the oscilloscope horizontal deflection. The discussion which follows will refer to photographs of the oscilloscope under different conditions. By way of preparation, it may be stated that the pulse transmitted through the small hole in the end plate of the waveguide will excite a large number of modes. There are, at 9,000 mc, approximately 40 modes which can propagate in this waveguide. Also, the coupling through the holes is so weak and the energy lost due to dissipation in the shorting plates is so small as to represent an attenuation which is small compared with the theoretical wall loss in the 500-foot long line. Therefore, as the pulse shuttles back and forth in the line, it will decay as though it had travelled on a straight long section of waveguide made up of 500-foot long segments identical to the single 500-foot section actually constructed.

Fig. 11 shows a photograph of the oscilloscope displaying the time interval immediately following the transmitted pulse. The pulse at the extreme left represents the transmitted pulse which passes directly from the transmitter hole to the receiver hole on the end plate of the waveguide. The blank time interval immediately following the transmitted pulse is about one microsecond long and represents the time of travel of energy down to the far end of the 500-foot line and back to the sending end. During this interval no pulses were received because the joints in the line produce little reflection. The first pulse after the transmitted pulse represents energy travelling in the mode which has the highest

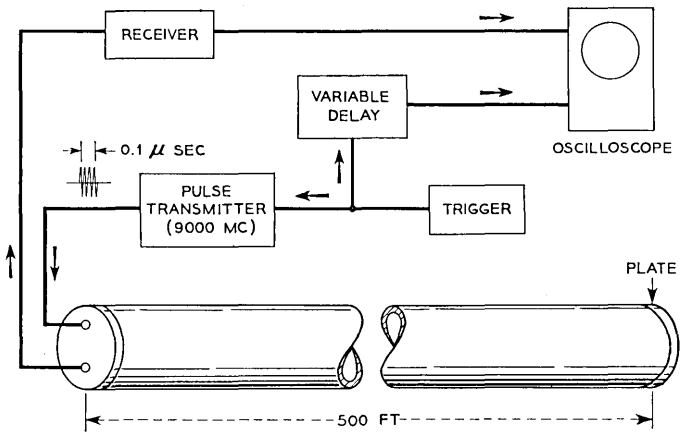

Fig. 10 — Diagram of equipment used for pulse tests of waveguide transmission.

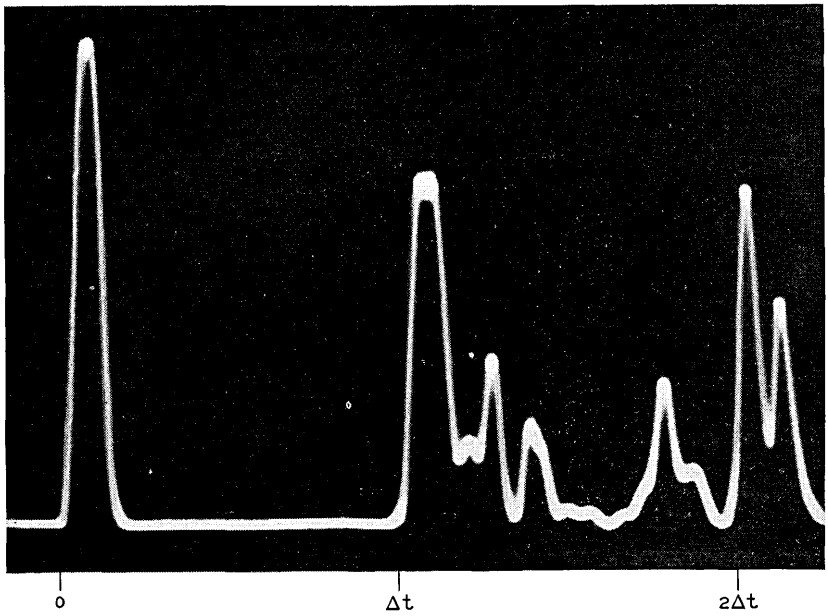

Fig. 11 — Photograph of cathode-ray tube presentation during the time interval immediately following the transmitted pulse.

group velocity. Pulses immediately following this first received pulse represent energy travelling in other modes whose velocities are lower and which therefore require more time for the one round trip of travel. At the time $2\Delta t$, we begin to observe pulses which have made two round trips in the line. If the transmitted pulse width were short enough, we could theoretically identify the mode in which the energy travelled by observing the time of arrival, since the velocities of propagation and the distance are known parameters. The $\frac{1}{10}$ microsecond pulse used in these experiments is not short enough to allow this kind of resolution on an individual mode basis. Something on the order of five or six modes have velocities so nearly the same that they cannot be resolved as separate pulses with the $\frac{1}{10}$ microsecond pulse after a single round trip in the 500-foot line.

Fig. 12 represents the same condition as Fig. 11, except that the horizontal time base has been changed to display the interval 0 to $14\Delta t$ instead of the interval 0 to $2\Delta t$. Fig. 12 shows fewer pulses in the time interval $6\Delta t$ to $12\Delta t$ than in the time interval 0 to $6\Delta t$. This is because energy travelling in some modes is attenuated more rapidly than that in other modes. For time delays greater than $10\Delta t$ the received pulses

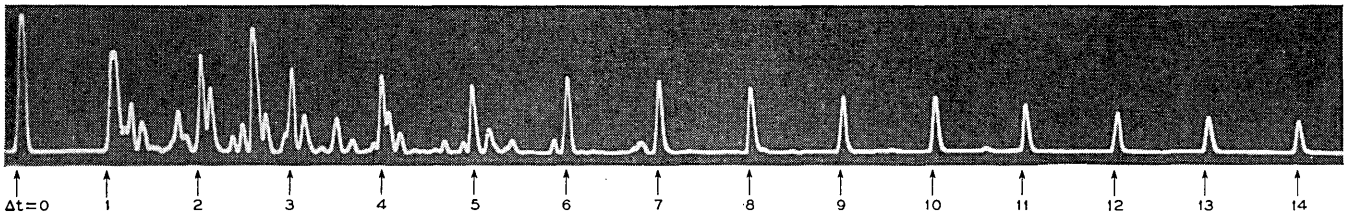

Fig. 12 — Same as Fig. 11 except that a longer time interval is shown.

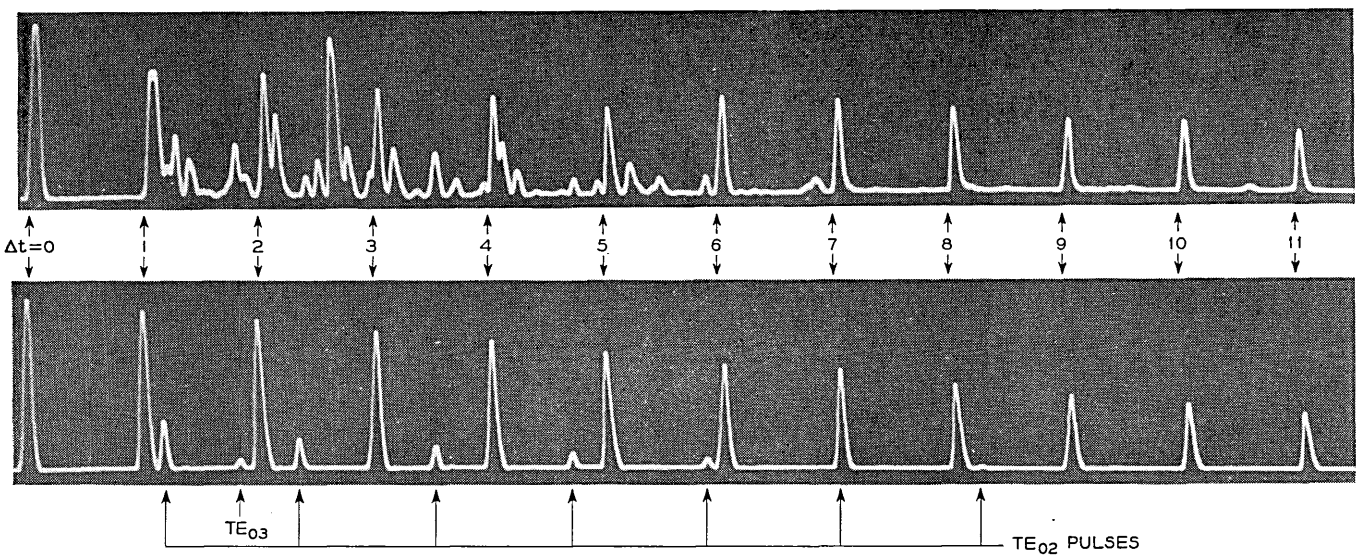

Fig. 14 — The effect of a mode filter in the 500-foot line. The unlabeled pulses in the lower part are in the TE_{01} mode.

appear at regular intervals and with smoothly decaying amplitude. This behavior indicates that the major portion of the energy in the line was travelling in a single mode, and we deduced that this mode was TE_{01} as follows: We observed that the velocity of propagation was near that for the TE_{01} mode by measuring the absolute time between pulses, averaged over many round trips. This excluded all but about 6 modes whose cut-off frequencies are near that of TE_{01} . Measurement of transmission loss was made by observing the rate of decay of the received pulses averaged over 10 or more round trips. The measured loss was found to be approximately 3 db per mile compared to a theoretical value of 1.9 db per mile for TE_{01} propagation. It follows that propagation must have

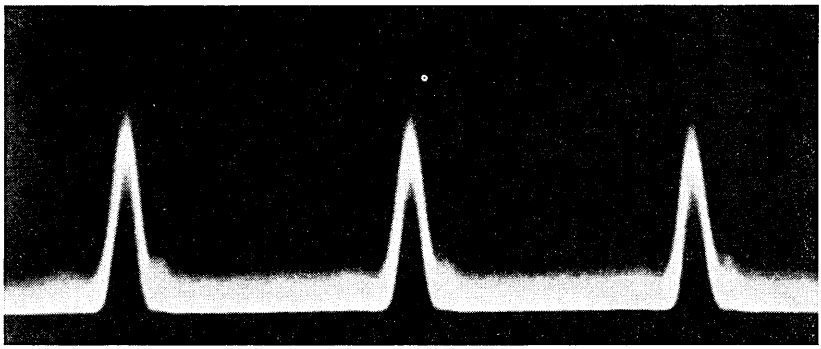

Fig. 13 — Record of pulses after 40 miles of repeated traversal over the 500-foot line.

been taking place in the TE_{01} mode, for all other modes near TE_{01} in velocity have theoretical losses well in excess of the observed value.

To summarize the effects shown in Fig. 12, a great many modes including TE_{01} were launched by exciting the waveguide through a small aperture in the end plate. All these modes propagated back and forth in the line for a while, but due to the fact that TE_{01} has appreciably less loss than the other modes, the energy remaining in the line after a suitable time delay was substantially all in the TE_{01} mode. This permitted measuring TE_{01} loss over a distance of many miles by allowing the energy to traverse the 500-foot line many times.

Fig. 13 records three successive trips of a pulse which had travelled up and down the 500-foot waveguide for a total distance of 40 miles. The pulse shape was still essentially the same as that of the transmitted pulse, although background noise had become clearly visible. We cer-

tainly can conclude from this that circular electric wave transmission over great distances is possible.

The long waveguide line also provided a very convenient way of demonstrating additional multi-mode transmission effects. For example, in a multimode medium we may use mode filters. One such filter may have a very low loss for the circular electric waves but very high loss to other waves. Such mode filters have been built, and Fig. 14 shows the transmission changes which result when introducing one. The upper half of Fig. 14 shows the photograph of the oscilloscope in the time interval 0 to $11\Delta t$ with no mode filter in the line. (This is the same as Fig. 12.) On introducing the mode filter into the waveguide, the received signal is altered as shown in the lower half of Fig. 14. We observed that energy propagating in the undesired modes has largely disappeared; it was absorbed by the mode filter.

There are still a few small pulses in the lower half of Fig. 14 which cannot be in the TE_{01} mode because of their time position. Starting at time $1.15\Delta t$, there is a series of regularly spaced pulses in Fig. 14 labeled TE_{02}, and a single small pulse labeled TE_{03}. The geometric placement of resistive material in the mode filter leads us to anticipate low filter losses for the entire circular electric (TE_{0n}) family of modes, and therefore the extra pulses were suspected of being in higher-order circular electric modes. Only two such modes, TE_{02} and TE_{03}, were above cutoff. The TE_{03} pulse was tentatively identified by noting that its group velocity was 55 to 60 per cent of that of the TE_{01} pulses. High attenuation in the TE_{03} mode prevented additional TE_{03} pulses from being observed.

In the case of the TE_{02} series of pulses, it was possible to get a fairly accurate measure of relative group velocity, confirming the identification as TE_{02}. Note that the seventh TE_{01} pulse coincides with the sixth TE_{02} pulse, and that the pulse at $7\Delta t$ shows on the TE_{01} train as being too large in amplitude.

The smooth decay of the TE_{01} train in the lower half of Fig. 14 in the interval Δt to $6\Delta t$ is in marked contrast to the corresponding pulse train in the upper half of the figure, and is graphic illustration of the important effects that mode filtering can produce.

Another very important transmission observation appeared during Mr. Beck's experiments with the 5″ diameter line. He observed that the attenuation, as measured by the amplitude decay of the shuttling pulse, was a function of the position of the piston at the far end of the line. Translations of the far end piston on the order of 10 to 40 centimeters changed the overall transmission from a condition in which the original

pulse shape was preserved for as many as 40 miles of travel (Fig. 13), to a condition wherein the shape of the pulse was badly distorted after only 3 or 4 miles of travel. This general behavior is illustrated by the photographs shown in Fig. 15. We will concentrate for the moment on the top two rows of photographs which record the pulse transmission in the bare waveguide as a function of distance of pulse travel for both favorable and unfavorable piston settings. However, all of the rows of the photographs were taken under such conditions as to permit direct comparison.

The photographs at the extreme left end represent the outgoing pulse and the first echo pulses which travelled one round trip in the line, approximately 340 yards. All the other photographs show two principal pulses which record two successive trips of the pulse as it passed the transmitting end (Fig. 10). The second photograph from the left represents the pulse as it passed the sending end after 10 and 11 round trips. The third picture from the left records the 20th and 21st trip, the fourth picture the 30th and 31st trip, etc. The numbers placed directly beneath the individual photographs represent the relative sensitivity of the receiver for that particular photograph. Reference receiver sensitivity was taken as the condition under which the 10th and 11th trip in the bare waveguide were recorded with a favorable piston setting (0 db beneath the photograph), and the designation -6 db under the adjacent photograph indicates that 6 db more receiver sensitivity was used in the latter case. The relation between display amplitude and actual pulse amplitude was approximately square law. The distance of pulse travel associated with each of the pictures is given at the bottom of the figure.

Comparing the top two rows representing favorable and unfavorable piston positions, we note that the attenuation was appreciably different — the values being 2.6 db per mile and 3.1 db per mile respectively. Serious distortion of the transmitted pulse also occurred for the unfavorable piston setting. Since the receiver in the experiment was sensitive to very many modes, one might suspect that the spurious pulses which appeared at more than 7,000 yards with the unfavorable piston setting might represent energy present in some of the other modes. Actually, all of the pulses and wiggles shown in the photographs at ranges greater than 3,500 yards were in the circular electric mode. This was deduced by first noting that every two successive pulses were not appreciably different from each other (see Fig. 15). If some of the distortion effects shown by the received pulse were due to energy being received in modes other than the signal mode, then successive pulses would be different in shape because the various modes have different phase constants. The very gradual change in pulse shape which did

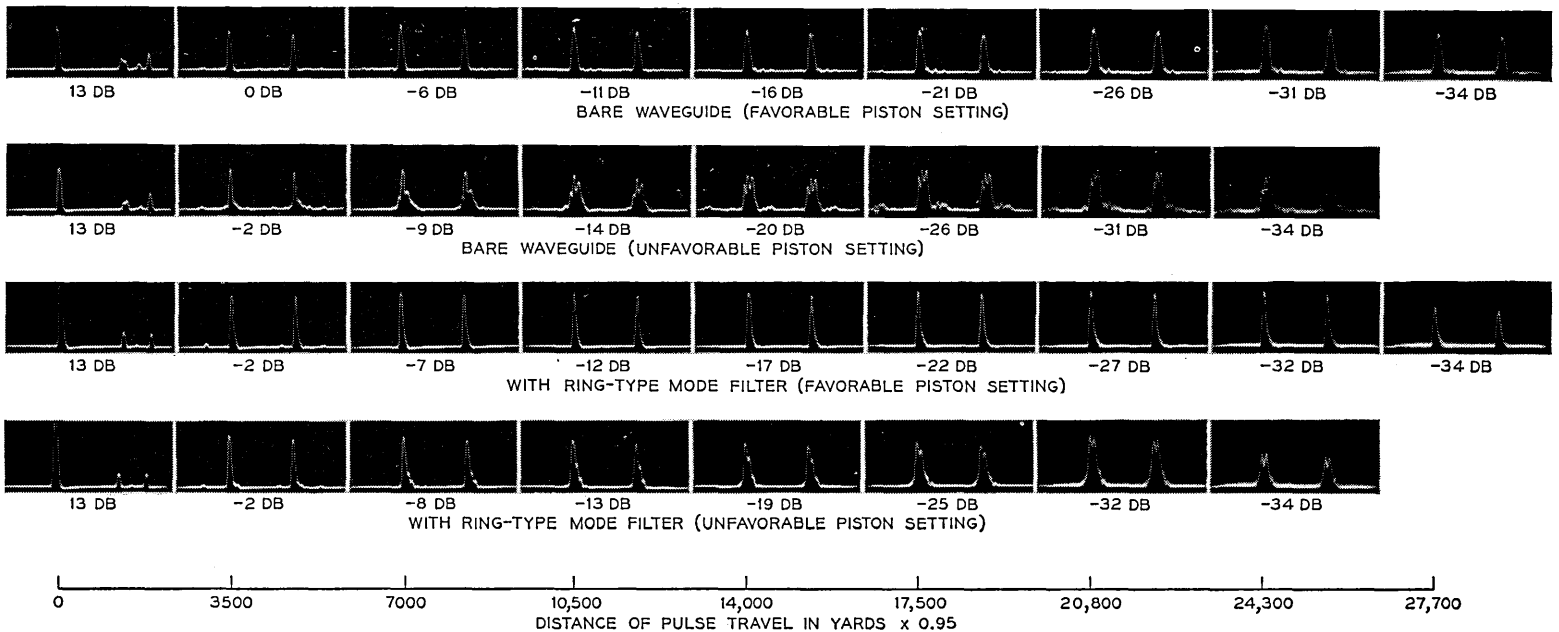

Fig. 15 — Photographs of pulse transmission with both favorable and unfavorable piston settings at the end of the 500-ft. waveguide.

occur as the pulse travelled up and down the line had the general form of an amplitude component which led or lagged the signal pulse by a constant interval but which gradually increased in amplitude with increasing distance that the signal pulse had travelled. Note, for example, the growth of spurious peaks before and after the signal pulse in row 2 of Fig. 15 at 3,500, 7,000, 10,500 and 14,000 yards of travel.

The explanation of this behavior involves transfer of energy from the circular electric mode to one or more of the unused modes of propagation and reconversion of the same energy back to the circular electric mode. This process is one of the characteristic features of multimode waveguide systems and is discussed at greater length in the following sections.

MODE CONVERSION AND RECONVERSION AS A SIGNAL-LOSS EFFECT

In beginning discussion of the mode conversion-reconversion phenomena, we take an idealized case of a dissipationless waveguide containing two deformities, Fig. 16. We assume a *c-w* signal entirely in the circular electric mode entering this waveguide. After passing the first deformity there will be energy present in some other mode, and this is designated as TX_1. When the combination of $TE_{01} + TX_1$ strikes the second deformity, another conversion takes place and the output will be a large TE_{01} component, two smaller components in the unused mode, TX_1 and TX_2, and a still smaller circular electric wave component, TE_{01}', which is due to reconversion of energy from TX_1 to the circular electric wave in traversing the second deformity. It can be shown[2] that for the proper distance between two identical symmetrical deformities, the wave emerging from the line may be purely circular electric; the two components TX_1 and TX_2 cancel each other under this condition. Another separation between the deformities results in a maximum energy transfer from circular electric to other modes. Therefore, it follows that any mechanism which varies the effective spacing between conversion points will produce conversion loss variations. This accounts for the change in the attenuation of the circular electric wave pulses in Fig. 15 as a function of the far end piston setting.

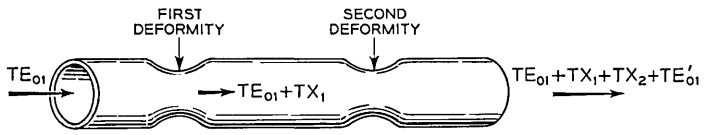

Fig. 16 — A distorted waveguide and the associated mode-conversion signal-loss effects.

MODE CONVERSION AND RECONVERSION AS AN INTERFERENCE EFFECT

The mode conversion-reconversion phenomena can also produce a signal distortion or interference effect. Fig. 17 shows another idealized waveguide containing two deformities, but in this case we assume a *short pulse* in the circular electric wave as the input signal. The amplitude-time plots below the waveguide show the energy present in the circular electric and unused waves respectively at various physical points along the line. The key item in this diagram is the time displacement between the converted energy TX and the circular electric wave energy at the input to the second deformity. This time difference appears as a result of propagation over an identical line length at two different group velocities which, in general, the circular electric and unused modes will possess. Since the signal and unused mode pulses strike the second deformity at different times the second conversion process results in energy appearing back in the circular electric wave at a time separated from the signal pulse itself. When the distance between deformities is too short for the pulses to be resolved at the second deformation, the result will be a distortion of the signal pulse rather than the appearance of a separate pulse.

The above very much simplified picture of the mode conversion and reconversion effects allows one to visualize several general properties of this phenomenon:

1. In general, there will be a large number of unused modes which will be coupled to the signal mode through the various imperfections in the transmission line. Since these unused modes have unequal phase constants, and since the imperfections will be randomly spaced along the line, the reconverted signal pulses in a time-division system will be spread

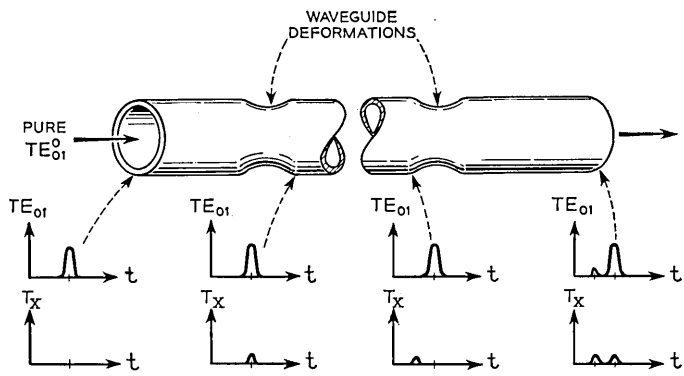

Fig. 17 — Signal interference effects due to mode conversion and reconversion.

out on the time scale rather than appear as a single well-defined distortion pulse.

2. Because some of the unused modes of propagation have group velocities greater than the group velocity of the circular electric wave, the reconverted energy pulses may reach the receiving end of the transmission system before the signal pulse itself appears.

3. The reconverted energy will, in general, be out of phase with the signal from which it was derived. When the differential time of travel in the unused mode is short, the principal effect of the conversion-reconversion process will be to distort the signal wave. When the differential time of travel in the unused mode becomes as large as the reciprocal of the modulation frequencies involved, the reconverted energy will appear more like an echo. In a time-division system such echo pulses would interfere with the pulses representing other signal components. Because of the large number of conversions contributing at random time delays, this "echo-interference" may be unintelligible. In this sense the interference may be thought of as a noise effect, just as multi-channel cross-talk due to amplitude non-linearities in a single sideband AM system may be thought of as noise.

4. The general case of a signal pulse, both preceded and succeeded by a series of reconverted energy pulses, is sketched in Fig. 18. It is quite apparent that if the reconverted signal pulses are allowed to become of the same order as the signal pulse itself, even a pulse code modulation system will be rendered inoperative. Other types of modulation will experience difficulty at appreciably smaller magnitudes of reconverted energy.

5. The level of the reconverted energy relative to the signal is determined by the transmission medium. It is not possible to avoid this interference by using more power at the sending end of the transmission link, for the interference rises with the signal. The need for low-noise receivers is just as acute as in other transmission systems, because better noise figure means that correspondingly less power is required from the transmitter.

6. If the loss to the mode TX is very large in the region between successive waveguide imperfections, the TX pulse can be attenuated to a negligibly small value before reaching the second deformation, thus preventing any significant reconversion back to the signal mode.

7. Limits can be placed on the time delay between the signal energy and the reconverted energy returned to the signal mode. The lower limit on this time delay is obviously zero, corresponding to a series of imperfections very close together. The upper limit can be taken as the differ

ence between the times it takes the signal mode and the unused mode to travel a distance in which there is a large difference between the ohmic losses in the unused-mode and signal mode. This concept may be expanded as follows: Energy is transferred at an imperfection located at coordinate z_1 on the axis of propagation, producing an amplitude in unused mode x. At z_2 on the axis of propagation the amplitude in mode x will be attenuated relative to the signal-mode amplitude at the same point by the factor

$$\epsilon^{-(\alpha_x-\alpha_1)(z_2-z_1)}$$

where α_x and α_1 are the normal heat loss coefficients in mode x and the signal mode respectively. When the exponential factor is small enough (i.e., z_2 large enough), reconversion will no longer be important compared to reconversion near z_1. For order of magnitude we might assume that 10 db more attenuation for the x-mode amplitude than for the signal-mode amplitude would render further reconversion unimportant.

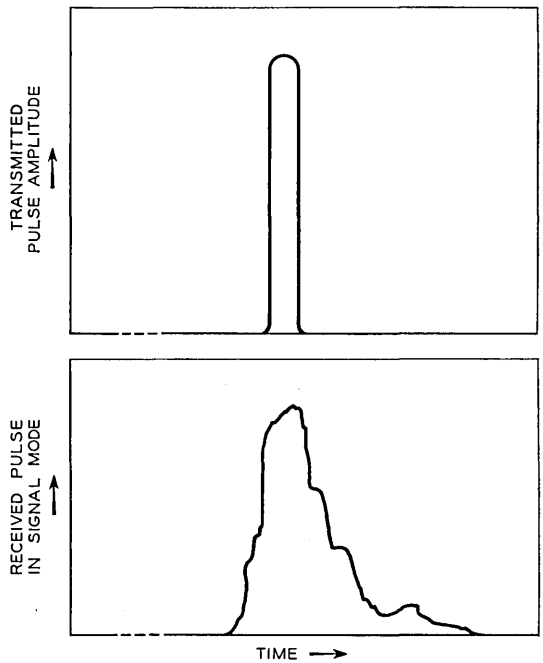

Fig. 18 — Schematic of signal distortion due to conversion and reconversion effects in a line with randomly placed conversion points.

Then we know the distance $(z_2 - z_1)$ from

$$(z_2 - z_1) = \frac{1.15}{(\alpha_x - \alpha_1)}.$$

The corresponding "upper limit" on time delay between the signal and the reconverted energy in the signal mode is

$$t = (z_2 - z_1)\left(\frac{1}{v_x} - \frac{1}{v_s}\right) \tag{1}$$

where v_x and v_s are the group velocities in the mode x and the signal mode.

It is well known that the unused modes of a circular electric waveguide have attenuation coefficients which are appreciably larger than attenuation coefficients for the circular electric wave itself. In the light of this attenuation to the unused modes plus the fact that the reconverted energy has undergone two mode conversion losses before it reaches the signal mode again, one might wonder whether the mode conversion-reconversion phenomenon would really be an important effect. The first indication that the magnitudes of the reconversion amplitudes are significant came during experimental work on the 5″ diameter 9,000-mc line, described above in connection with Fig. 15. A more quantitative theoretical discussion which follows shows that the reconversion phenomena will continue to be important even when mode filters are introduced into the line.

The effects of the mode conversion-reconversion process are very similar to the effects of multipath transmission through the atmosphere. In microwave radio there are under unusual fading conditions 2 or 3 subsidiary signals, and these are representative of propagation over different path lengths in space but at the same velocity of propagation. In the waveguide, there will in general be a large number of subsidiary transmission paths, each of which corresponds to the identical distance of propagation but at velocities of propagation which are different for the various modes. The radio multipath phenomenon exists only occasionally, whereas the waveguide multipath phenomenon is a steady characteristic present 100 per cent of the time. Long-distance radio transmission in the 6–20 mc region by way of the ionosphere encounters multi-path effects more like those expected in the waveguide, with the exception that waveguide multi-path effects are expected to have far greater short-time stability.

Quantitative relations describing the conversion-reconversion process may be derived by considering an infinitely long waveguide composed of

a series of identical sections each containing a single mode-conversion discontinuity at the midpoint. (The actual 500-foot line contains many conversion points, as will be discussed later, and is effectively repeated over and over as the pulse traverses the identical line many times.) A very short signal-pulse of unit amplitude is assumed as an input to the idealized line containing identical conversion points. After the first conversion point the amplitude in the unused mode is k and the amplitude in the signal mode is $(1 - k^2)^{1/2}$, where k is a measure of the size of the conversion irregularity. There is no reconverted wave at this point since there is no input to the first conversion point in the unused mode. After the second conversion point, however, there is a reconverted-wave amplitude $k^2\epsilon^{\theta_x}$, where θ_x is defined below. After the n^{th} conversion point, it may be shown that the amplitude in the signal mode is

$$\epsilon^{(n-1)\theta_1}(1 - k^2)^{n/2} \tag{2}$$

the amplitude in the unused mode is

$$\begin{aligned} &k(1 - k^2)^{\frac{(n-1)}{2}}\epsilon^{(n-1)\theta_x} \\ &+ k(1 - k^2)^{\frac{(n-1)}{2}}\epsilon^{(n-2)\theta_x+\theta_1} \\ &+ k(1 - k^2)^{\frac{(n-1)}{2}}\epsilon^{(n-3)\theta_x+2\theta_1} \\ &+ \text{etc. to} + k(1 - k^2)^{\frac{(n-1)}{2}}\epsilon^{(n-1)\theta_1} \end{aligned} \tag{3}$$

and the reconverted wave amplitude is

$$\begin{aligned} &(n - 1)k^2(1 - k^2)^{\frac{(n-2)}{2}}\epsilon^{\theta_x+(n-2)\theta_1} \\ &+ (n - 2)k^2(1 - k^2)^{\frac{(n-2)}{2}}\epsilon^{2\theta_x+(n-3)\theta_1} \\ &+ (n - 3)k^2(1 - k^2)^{\frac{(n-2)}{2}}\epsilon^{3\theta_x+(n-4)\theta_1} \\ &+ \text{etc. to } k^2(1 - k^2)^{\frac{(n-2)}{2}}\epsilon^{(n-1)\theta_x} \end{aligned} \tag{4}$$

in which

z = distance between adjacent conversion points

α_1 and α_x are the heat loss coefficients (applying to wave amplitudes) for the signal and unused modes respectively.

β_1 and β_x are the phase constants for the signal and unused modes respectively.

$\theta_1 = -(\alpha_1 + j\beta_1)z$

$\theta_x = -(\alpha_x + j\beta_x)z$

k = amplitude of the converted wave for a single conversion point and for unit wave incident on the conversion point.

In deriving the series represented by (2), (3) and (4), it is assumed that all of the converted power travels in the forward direction, and that reflection effects are negligible. These conditions are typical of imperfections in practical multi-mode waveguides.

When the input pulse for the idealized line is sufficiently short, the various terms of (3) and (4) (representing successive conversions and reconversions) are non-overlapping pulses. It is instructive to write down the ratio of the signal-wave amplitude to the reconverted wave amplitude which is separated from the amplitude of the signal wave at the same point by the time difference $z(1/v_x - 1/v_1)$, in which v_x and v_1 are group velocities. This ratio is [ratio of (2) to the first term of (4)]

$$\frac{(1 - k^2)}{(n - 1)k^2\epsilon^{-(\alpha_x - \alpha_1)z}} \tag{5}$$

It is clear from this ratio that the signal wave may be smaller than the reconverted wave if n is sufficiently large. Physically, what happens is that the reconverted amplitude created at each successive conversion point adds in phase with the reconverted wave amplitude present at that point due to previous conversions and reconversions. This happens of course because the line contains identically spaced conversion points. With random location of conversion points, a less severe build-up of reconverted wave energy would certainly occur.

The first and second reconverted pulses [i.e., the first and second terms in (4)] are separated by the time difference $z(1/v_x - 1/v_1)$ and the ratio of the amplitude of the second to the first reconverted pulse is

$$\frac{(n - 2)}{(n - 1)}\epsilon^{-(\alpha_x - \alpha_1)z} \tag{6}$$

The largest amplitude existing in the unused mode at a point immediately following the n^{th} conversion is the component converted at the n^{th} conversion point and the ratio of this component to the amplitude of the signal pulse after the n^{th} conversion is:

$$\frac{k}{(1 - k^2)^{1/2}} \tag{7}$$

Note that this ratio is independent of n. The unused-mode amplitude converted at the first conversion point is, after n trips and relative to

the signal pulse at the same point,

$$\frac{k}{(1 - k^2)^{1/2}} \epsilon^{-(n-1)(\alpha_x - \alpha_1)z} \tag{8}$$

These expressions show that the largest unused mode amplitude at any point in the line will be the one most recently converted from the signal wave and will be nearest to the signal pulse in time position.

The sketch shown in Fig. 19 represents schematically the signal pulse, the unused mode pulse and the reconverted pulse amplitudes after n trips past the conversion point. It is interesting to note that the most recent (the n^{th}) conversion to the unused mode appears in a time position close to the signal pulse whereas the most recent reconversion appears at a time far removed from the signal pulse.

Let us investigate the ratios (5) through (8) under conditions representative of those in the 5″ diameter waveguide line. Row 2 of Figure 15 shows that after 40 trips down and back on the 500-foot line, i.e., after 80 trips past the center of this line (where there is assumed a single conversion point), the amplitude of the reconverted pulses which appear just before and just after the signal pulse are about equal to the signal

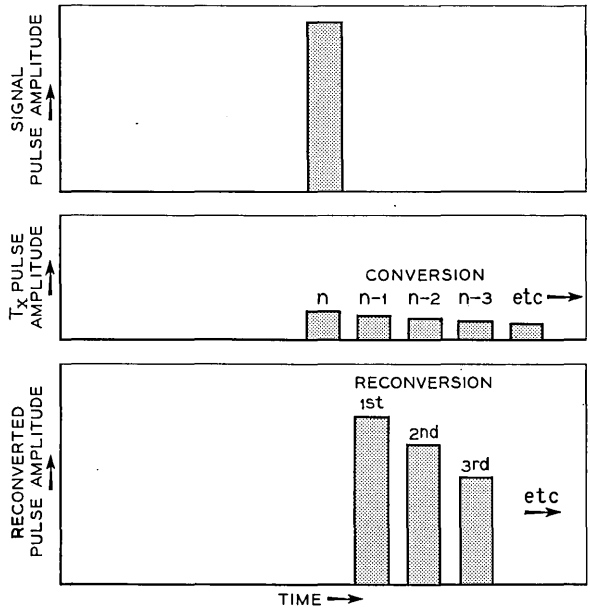

Fig. 19 — Sketch of pulse amplitudes in a line having equally spaced conversion points.

pulse. We may then set the ratio of signal-wave amplitude to the reconverted-wave amplitude, as given by (5), equal to unity at $n = 80$. We know the theoretical values of the heat loss coefficients α_x and α_1 for the unused and signal modes respectively. Allowing modest increase for surface roughness effects, we will assume the heat loss for the signal (TE_{01}) wave to be 0.2 db between conversion points (about 500 feet apart) and the loss for the unused mode to be about five times this or 1 db between conversion points. Substituting these numbers into (5) we can solve for the conversion coefficient k which is necessary to obtain the observed equality between reconversion and signal wave amplitudes at the $n = 80$ point. Such a calculation yields a value of k of 0.117 and this implies a conversion loss which is 32 per cent of the heat loss for the signal wave. Direct observations of mode conversion (to be described) show that the conversion losses in the 5″ diameter line must be at least as large as this. Therefore, we have confirmed that the basic mechanism which we are discussing can indeed account for a break-up of the signal pulse of the general type actually observed.

The calculated ratio of the second reconverted pulse compared to the first reconverted pulse from (6) for the $n = 80$ condition described above is only -0.5 db, which shows that if we really had a single conversion point which could add exactly in-phase in the manner assumed, we would have in our photographs an even worse series of reconverted pulses than actually do appear.

Again for the $n = 80$ condition in the 5″ line the calculated amplitude of the unused mode component relative to the signal component at the same point is -18.5 db for the n^{th} conversion and, with the square law response of the display system, it is to be expected that such an amplitude would not be observed. This fits satisfactorily with the observation that the pulses in the 5″ line, after 40 round trips, are all in the circular electric mode even though the conversion-reconversion process is significant.

As already implied, the mode conversion situation for the actual 500-foot line, which is represented by the photographs of Fig. 15, is more complicated than the idealized line just analyzed. First, note that in the case of the nonoverlapping pulses in the idealized line the position of the far end piston (i.e., the spacing between the conversion points) has no critical effect on the conversion and reconversion amplitudes. In the experiment (Fig. 15), far-end piston movements of a foot or so caused very significant changes in the observed conversion and reconversion. This is a consequence of the fact that in the physical 500-foot line, there were a large number of conversion points, and the pulse width

was not small enough to produce non-overlapping effects. If all of the conversions in the 500-foot length were between TE_{01} and a single additional mode, then it may be shown that the piston position which would cause in-phase addition of components from one conversion point on successive traversals of the 500-foot line would also cause in-phase addition for the components from all of the other conversion points. However, in the physical line significant conversion takes place between TE_{01} and three or more other modes, each of which has a different phase constant and each of which requires a different piston setting for in-phase (or out-of-phase) addition of components generated on successive traversals. Whereas it is not possible to relate the shuttle-pulse observations of Fig. 15 to a simple quantitative theory, the general behavior of the experimental line is in qualitative agreement with the conversion-reconversion explanation.

The favorable piston setting in Fig. 15 represents an infinite line in which approximate amplitude cancellation of mode conversion effects is achieved, whereas the unfavorable piston setting corresponds to an infinite line in which approximate amplitude addition of mode conversion effects is experienced. Since the former is optimistic and the latter is pessimistic compared to what might be expected in a physical long line, it has been our practice to average the loss data obtained at the favorable and unfavorable piston positions.

The expression (5) for the ratio of the signal pulse to the reconversion pulse shows that appreciable loss between the conversion points, represented by the factor $\epsilon^{-(\alpha_x-\alpha_1)z}$, is an effective way of reducing the reconverted wave effects. During the early experiments it was found that mode filters did reduce the influence of the far end piston. In Fig. 15, rows 3 and 4, show records of the 5″ line pulse transmission with a single mode filter at the sending end. The mode filter did not completely eliminate the phasing effects of the piston, and this may be expected for at least two reasons: (1) the mode filter itself may cause some mode conversion, and this mode conversion component will be influenced by the piston setting, and (2) the attenuation introduced by the mode filter is not sufficient to completely eliminate the wave components in the unused mode.

Even in the presence of the best mode filter now available, row 4 of Fig. 15 shows that the conversion-reconversion phenomenon does take place. The reconverted pulse amplitudes show up at 7,000 yards in row 4 as small distortions on the right hand side of the signal pulse, and these distortions grow in magnitude as we proceed to the right in the series of pictures representing more trips past the conversion points.

In the 5″ line there are several modes for which the delay factor $z(1/v_x - 1/v_1)$ does not result in a separation between the reconverted pulse and the signal pulse until z is on the order of 5,000 feet. Thus, we might expect the conversion-reconversion phenomenon to broaden the signal pulse. This does indeed take place even for the favorable piston setting of the bare waveguide, as shown by the top line of pulses in Fig. 15. The pulses at the distance 3,500 yards are sharper than those pulses at the distance 27,700 yards. However, addition of the mode filter (which introduces negligible signal attenuation) does appreciably sharpen the pulse at the 27,700 yard distance (row 3).

Thus, on the basis of pulse transmission observations on the 5″ line and a simple theoretical analysis, we conclude that the conversion-reconversion phenomenon will be important in a waveguide system, and that it is important to have as much dissipation as possible present in the unused modes of propagation.

ANALYSIS FOR CONTINUOUS MODE CONVERSION

The traveling pulse type of theoretical analysis utilized in the preceding section can be extended to describe a more realistic spatial distribution of conversion points and to include a series of unused modes instead of only one. An extension of this type is required in order to calculate directly the behavior which might be expected in a waveguide composed of randomly disposed irregularities.

A much simpler mathematical treatment, originally suggested to the writer by J. R. Pierce, is to assume uniform mode conversion along the direction of propagation and to represent this condition by a differential equation. An analysis of this type is attached as an appendix. The work includes the assumption of quadrature addition of conversion and reconversion components, and the total magnitude of such components given by the analysis may be thought of as the rms average of the conversion magnitudes in waveguide lines containing randomly located imperfections. Any single line might show somewhat more or less conversion effects, a factor of ± 10 db probably being adequate to cover most lines containing randomly located imperfections. If practical lines show appreciable correlation between the spacings of the conversion points, the reconverted-wave magnitude would become greater. The analysis has the advantage of being simple and understandable and should give overall trends accurately.

This analysis shows that the waveguide performance with regard to conversion-reconversion effects is completely specified with knowledge

of (1) the ratio of the conversion coefficient (a_{1x}) to the true dissipation coefficient* in the signal mode (a_{1h}), (2) the ratio of the heat-loss coefficient of the unused-mode to that for the signal-mode (a_{xh}/a_{1h}) , and (3) the length of the transmission line specified in terms of decibels of heat loss to the signal wave. For a given ratio of conversion-loss to heat-loss, the same ratio of signal-power to reconverted-wave power will be present for a long low-loss waveguide as for a short high-loss waveguide. This makes it important to determine the ratio of conversion-loss to heat-loss for waveguides of several nominal attenuation coefficients and to predict these effects theoretically insofar as it is possible.

Another result of this analysis is plotted in Fig. 20, which shows the ratio of the signal power to the power in the unused mode at the end of the line, with transmission-line heat loss as the abscissa. These curves have been plotted for a fixed magnitude of conversion loss coefficient (a_{1x}) equal to 50 per cent of the true heat loss coefficient (a_{1h}) and for ratios (a_{xh}/a_{1h}), heat loss in the unused mode to heat loss in the signal mode, between 2 and 100. These values are typical of solid round waveguide without mode filters. It is interesting to note in Fig. 20 that the magnitude of the unused mode power relative to the signal mode power reaches very nearly a constant value in a transmission line length of only ½ to 1 db, except for extremely low ratios a_{xh}/a_{1h} . Physically what is happening is that the unused mode power becomes dissipated through heat loss about as rapidly as it is created by mode conversion, after an initial short transmission line length.

Fig. 21 shows the ratio of signal power to reconverted wave power as a function of transmission line length for the same conditions described in connection with Fig. 20. A heat loss ratio on the order of 2 to 10 is typical of important modes in solid round waveguide without the addition of mode filters,† and Fig. 21 shows that the ratio of signal-to-reconverted wave power for such a medium becomes poorer than 20 db for transmission line lengths longer than 1.5 to 2 db. Although there is some uncertainty as to the precise interpretation which may be placed on the signal power to reconverted wave power calculated in this manner, since the time relations in connection with a definite modulation method are not included, it seems evident that a solid copper tube without mode

* There is a very significant difference between the effects of signal power loss to other modes through conversion and signal power loss due to dissipation in the waveguide walls. However, it does not matter here whether the latter be due to surface roughness, chemical impurity or just the theoretical minimum heat loss for ideal copper. Therefore, all of the heat loss effects are combined into the single coefficient, a_{1h} .

† See the appendix for further discussion.

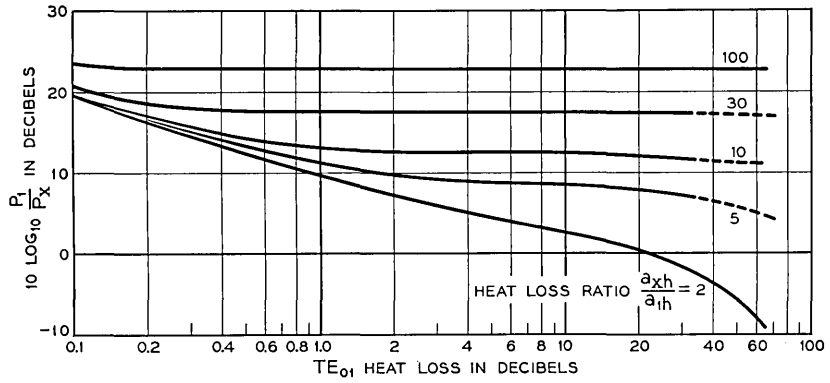

Fig. 20 — Ratio of TE_{01} signal power (P_1) to X-mode power (P_x) versus line length for conversion coefficients (a_{1x} and a_{x1}) equal to one-half the heat loss coefficient (a_{1h}).

filters is very unlikely to be satisfactory for long distances as a communication medium. However, the addition of mode filters will raise the heat loss to the undesired waves, and the latter improves the signal to reconverted-power ratio directly as the ratio of heat loss in the unused wave to heat loss in the signal wave. Thus, as shown on Figure 21, a heat loss ratio on the order of 500 would produce a ratio of signal power to reconverted wave power on the order of 20 db at a transmission line length of 60 db. It may be shown that the magnitude of the signal to reconverted wave power varies as the square of the conversion to heat loss coefficient ratio a_{1x}/a_{1h}.

The sharp break downward on the right hand end of the curves for $a_{xh}/a_{1h} = 2$ in Figs. 20 and 21 represents the condition wherein the power in the reconverted wave becomes comparable to the power remaining in the signal wave.

We next consider the improvement in signal fidelity which results from the introduction of ideal mode filters. We shall assume the ideal mode filters have a matched impedance for all modes, very high transmission loss to the unused modes, and no transmission loss for the signal mode. The improvement in the ratio of signal power to reconverted wave power due to the addition of such filters is shown in Fig. 22. This plot has been calculated for a total line length of 20 db heat loss, but the conclusions are valid for any line length wherein the signal wave power remains appreciably larger than the reconverted wave power. We observed that mode filters improve the signal-to-reconverted-wave powers very slowly when placed far apart, typical improvements ranging be-

tween 2 and 8 db for 1 db separation between mode filters. The larger improvement is obtained when the heat loss in the undesired mode is more nearly comparable to the heat loss in the signal mode before the addition of the mode filter. For very small spacing between the mode filters, addition of the ideal mode filters improves the signal-to-reconverted wave power by very large factors.

A somewhat different form of effect due to the addition of mode filters is observed if the line before the introduction of the mode filter has a reconverted wave power larger than the signal power. Under this condition, relatively large mode-filter spacings bring about a large improvement in signal-to-reconverted wave power, but this is due to the very poor condition present before filtering. It is doubtful whether the transmission line would become very useful without rather strong mode filtering of the type represented in Figure 22.

We may conclude that the mode filters should be placed very close together, preferably at spacings of less than .1 db heat loss to the signal wave. We may also conclude that the transmission of signals over distances corresponding to the order of 60 db heat loss will require either

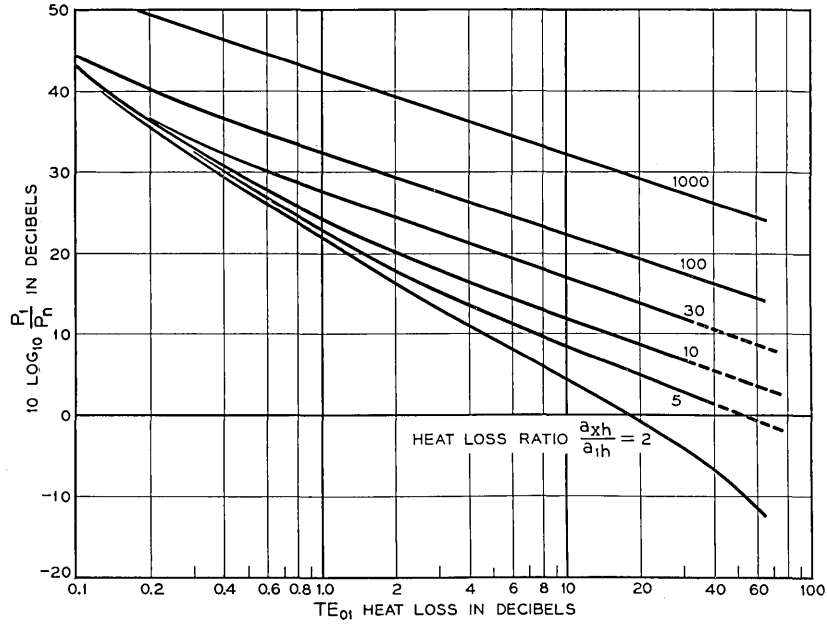

Fig. 21 — Ratio of TE_{01} signal power (P_1) to reconverted TE_{01} power (P_n) versus line length for $a_{1x} = 0.5\ a_{1h}$.

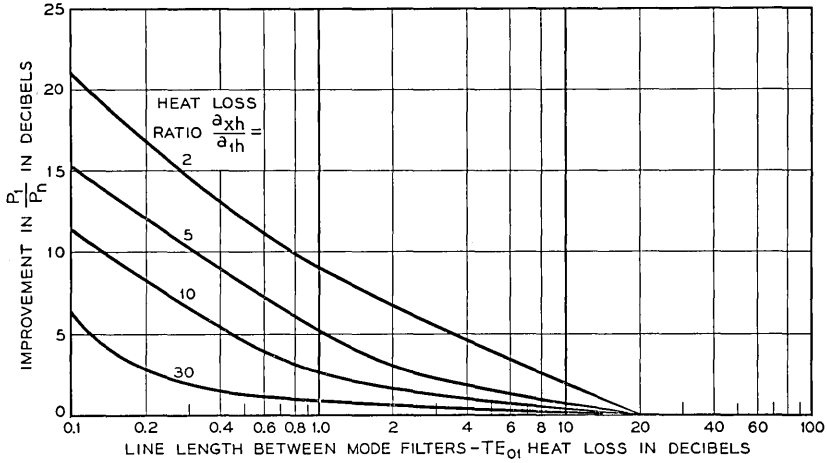

Fig. 22 — Improvement in P_1/P_n due to addition of mode filters. Total line length equals 20 db TE_{01} heat loss. (Plotted for $a_{1x} = 0.5\ a_{1h}$).

the spacing of mode filters at less than 0.1 db to the signal wave or use of a continuous form of transmission line having a ratio of heat loss in the unused mode to the heat loss in the used mode on the order of 500.

DIRECT EVALUATION OF MODE CONVERSION MAGNITUDES

The important influence that mode conversion effects are expected to exert on signal fidelity lead us to make direct evaluations of the conversion coefficients. The direct evaluation consisted of transmitting (actually or in imagination) a pure circular-electric wave into one end of a waveguide section and, by measurement or by calculation on the basis of known geometry, determining the relative magnitude of the power converted to the unused modes.

The simplest experimental technique for analyzing mode impurities consists of a short radial probe at the guide wall. The radial probe responds to energy in any mode of propagation except the circular electric family, and serves as a versatile instrument for measuring the *order of magnitude* of mode conversion effects. The limitations of the technique stem from (1) the fact that the probe responds to the vector sum of the amplitude of the radial electric field components of about 35 modes (in the 5″ line case), and this sum varies with circumferential and longitudinal position of the probe even though the power present in the modes is constant; and (2) the fact that the sensitivity of the probe response to a given magnitude of power in the guide is variable from mode to mode,

being (in an extreme case) 25 db greater for TE_{91} than for TE_{13} in the 5″ pipe. For the majority of modes, however, the latter variation is ±3 db, and the maximum probe response as a function of curcumferential and (to a limited extent) longitudinal position can be determined without excessive labor.

The probe technique of mode conversion evaluation was first applied by M. Aronoff to the individual sections of the 5″ diameter experimental line. He found that the average indication of conversion for the (approximately) 20 ft. lengths of pipe was 29.5 db below the signal wave power; since the power loss due to dissipation in the walls for a 20 ft. section is about 27 db below the signal wave power, the individual pipe measurements gave an order-of-magnitude estimate of 0.55 for the ratio of conversion loss to heat loss (a_{1x}/a_{1h}). Four mechanically distorted sections of line, previously considered satisfactory, were identified and discarded on the basis of this approach.

A. C. Beck and M. Aronoff next applied the probe technique to the 5″ diameter line assembled into lengths of 145 feet, 270 feet, and 500 feet. The indications of converted power were −17 db, −10.5 db, and −13 db respectively (at wavelengths near 3.2 cm) which is compatible with the hypothesis of random addition of a number of conversion components. Since the heat-loss power for the 500-foot line is about −13 db compared to the incident signal power, the 500-foot line radial probe measurement gave an order of magnitude estimate of 1.0 for a_{1x}/a_{1h}, in fair agreement with the value of 0.55 from single pipe measurements. The probe indication of conversion as a function of frequency for the 500-foot line is plotted in Fig. 23, which shows that quite a number of

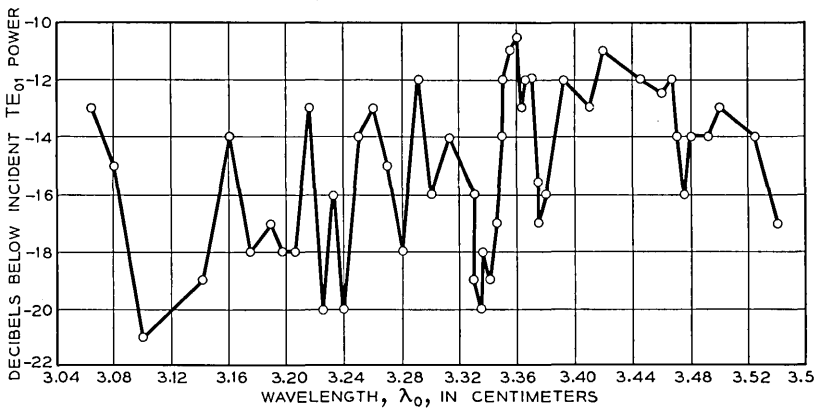

Fig. 23 — Probe recording of converted power in the 500-foot line.

conversion components were present; similar data were observed on the shorter lengths of line.

Azimuthal distributions of probe response showed conclusively that very little conversion was present to the 13 TE_{nm} and TM_{nm} modes having an index "n" of four or more.

A much more precise though more elaborate method of evaluating the conversion coefficients involves use of coupled wave transducers.[2] Such devices respond to only one mode and have known sensitivities, and therefore permit truly quantitative measurements. (At the present time this approach requires a separate transducer for each mode to be evaluated, a somewhat cumbersome procedure, but in principal the mode transducers can be made "tunable" for a series of modes.) A. C. Beck and M. Aronoff applied this more accurate method of conversion coefficient determination to several modes of the 500-foot line, and Table I shows the values averaged over the frequency band.

TABLE I — AVERAGE RATIO OF CONVERSION TO TE_{01} HEAT LOSS WITH TE_{01} EXCITATION

Mode	a_{1x}/a_{1h}
TE_{11}	0.21
TM_{11}	0.05
TE_{21}	0.14
TE_{31}	0.05
TM_{01}	<0.001
Total	0.47

The estimated absolute accuracy is between 10 per cent and 20 per cent for these ratios. The variation of the conversion coefficients as a function of frequency is shown in Fig. 24 for two of the important modes and for the total of the modes given in Table I. The total of the modes measured is consistent with the probe indications, though the accuracy of the latter is low enough that this should not be interpreted as proving there are no other important conversion contributors.

The above direct evaluation of mode conversion in the 500-foot line yielded magnitudes that are sufficient to explain the conversion-reconversion process as already outlined. We wished to extend our experience with this particular line to higher-frequency lines which might be built and to absolute tolerances that might be placed on the construction of a new line. Toward this end, theoretical relations were derived by S. P. Morgan, Jr., for the mode conversions to be expected due to waveguide ellipticity, and due to the tilt and offset which may be expected to occur at the junction of two sections. Experimental work was done by M.

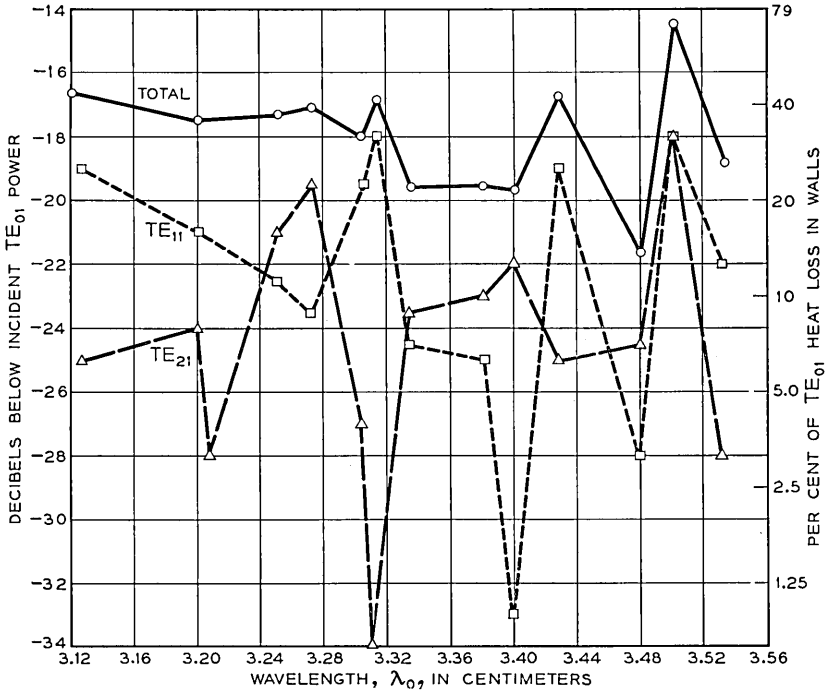

Fig. 24 — Observed conversion from TE_{01} to TE_{11}, TE_{21}, and the sum of TE_{11}, TE_{21}, TM_{11}, TM_{01}, and TE_{31} for the 500-foot line.

Aronoff[3] at 9000 mc on accurately created imperfections of the above types, and he found excellent agreement. We proceed to use the theory to compare the present line to a hypothetical 50,000 mc line. The results are given in Table II.

These computations represent a single oval section or a single tilted or offset joint. The converted power varies as the square of the tilt angle, as the square of the offset distance, and as the square of the difference between the major and minor diameters. The total ratio of converted power to heat loss power depends on the number of conversions per unit length. For the accuracy of constructing a 2″ diameter line assumed in Table II, the amount of mode conversion to be expected is not appreciably different from what appears attributable to known mechanical imperfections in the 500-foot line already discussed.

Another waveguide property of interest is the way the mode conversion magnitudes vary across the frequency band in a fixed pipe, and Table III shows this for the 2″ diameter line.

TABLE II — MODE CONVERSION COMPARISON OF 9000-MC 4.732″ DIAMETER LINE WITH A HYPOTHETICAL 50,000-MC 2″ DIAMETER LINE

Imperfection		Pipe Diameter	Frequency	Magnitude of Converted Power	
Type	Magnitude			Percentage	Mode
		inches	*Mc*		
Ovality	16* mils	4.732	9000	0.53%‡	TE_{31}
Ovality	4*	2.0	50,000	0.18%‡	TE_{31}
Tilt	1°	4.732	9000	0.34%	TE_{12}
Tilt	1°	2.0	50,000	2.0%	TE_{12}
Offset	10† mils	4.732	9000	0.008%	TE_{12}
Offset	2.5†	2.0	50,000	0.003%	TE_{12}

* Difference between major and minor diameters.
† Separation of guide axes.
‡ These are upper-limit values, based on the length of the oval section of pipe which would produce maximum mode conversion, and based on a cross-sectional shape (trifoil) which would produce the maximum of mode conversion.

It is interesting that two of the three conversion effects are essentially independent of frequency. Tilt at a waveguide junction introduces a phase-front error and would be expected to cause greater conversion effects at increasing frequencies. We shall see that bends produce a similar mode conversion, also due to a phase front error, that increases with increasing frequency.

THE BEND PROBLEM

The problem of transmitting the circular electric wave around bends was recognized as being important at an early date, and contributions to its solution were made by M. Jouguet,[4,5] W. J. Albersheim,[6] S. O. Rice,[7] and the writer.[8] The essence of the problem is as follows: A bend in a

TABLE III — FREQUENCY VARIATION OF MODE CONVERSIONS IN 2″ DIAMETER GUIDE

Imperfection		Mode	Magnitude of Converted Power		
Type	Magnitude		f = 24,000 mc	f = 50,000 mc	f = 75,000 mc
Ovality	4 mils	TE_{31}	0.22%*	0.18%*	0.20%*
Tilt	1°	TE_{12}	0.41%	2.0%	4.8%
Offset	2.5 mils	TE_{12}	0.003%	0.003%	0.003%

* These are upper-limit values, based on the length of the oval section of pipe which would produce maximum mode conversion, and based on a cross-sectional shape (trifoil) which would produce the maximum of mode conversion.

round guide causes coupling between the circular electric wave (TE_{01}) and the TM_{11} wave, and in round solid pipe the TE_{01} and TM_{11} waves are degenerate, i.e., they have identical phase constants. This degeneracy has the effect of bringing in phase all components transferred to TM_{11} no matter how gradually the bend might take place. Theory neglecting dissipation[4,7] shows that in a round pipe bent with any radius of curvature, the power will flow from the TE_{01} mode to the TM_{11} mode and back as a function of the total bend angle θ according to the relations

$$TE_{01} \text{ amplitude} = \cos\left(\frac{\pi}{2}\frac{\theta}{\theta_c}\right) \tag{9}$$

$$TM_{11}'' \text{ amplitude} = \sin\left(\frac{\pi}{2}\frac{\theta}{\theta_c}\right) \tag{10}$$

where

$$\theta_c = \frac{\pi}{2.32}\frac{\lambda_0}{a}$$

λ_0 = free-space wavelength

a = waveguide raduis

A solid round pipe is unsatisfactory for transmission of the circular electric wave around bends in the broad bands we seek to use.

One method of making the guide satisfactory in bends is to break the degeneracy between the TE_{01} and TM_{11} waves. Use of an elliptical pipe has been shown theoretically to be one way of doing this. For a 2″ diameter guide at 50,000 mc an eccentricity of about 0.3 permits a bending radius of 700 feet with theoretical bend losses in the range 0 to 0.17 db for any total bend angle; the heat loss coefficient for such an elliptic guide is about 35 per cent higher than for a perfectly round guide.[8]

Another method of avoiding bend losses is to introduce dissipation to the TM_{11} wave without adding loss to the TE_{01} wave. It has been shown[2] that a large difference between the attenuation coefficients of two coupled waves reduces the power transferred from the low-loss wave to the high-loss wave. Applied to the bend problem, this means that a structure with increased TM_{11} loss may be bent with less signal (TE_{01}) loss even though the phase constants might be degenerate. The reader is referred to the earlier paper[8] for a more complete discussion. There exist several alternate forms of circular electric waveguide (to be discussed) which have an attenuation coefficient for TM_{11} more than 5,000 times the attenuation coefficient for TE_{01}. The calculated extra loss in the bend region for such structures and for solid round pipe has been plotted

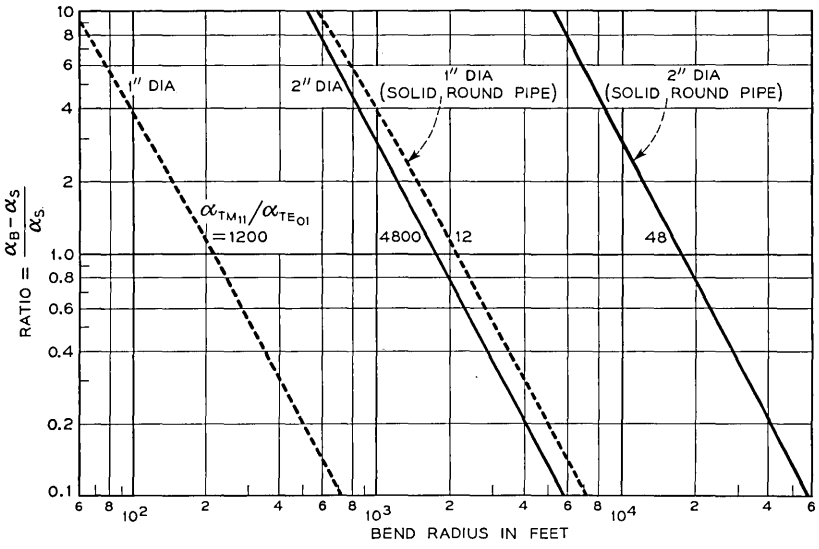

Fig. 25 — Computed change in the 50,000 mc TE_{01} attenuation coefficient due to bends in 1″ and 2″ diameter solid pipes and in modified guides having TM_{11} attenuation coefficients larger (than in solid pipe) by a factor of 100. α_B and α_s are the bend-region and straight-line attenuation coefficients, respectively.

as a function of bending radius in Fig. 25, assuming the TM_{11}-TE_{01} coupling due to the bend is the same in the altered guide as in solid round pipe. Whereas a bending radius of 17,500 feet causes a 100 per cent increase in TE_{01} heat loss for 50,000 mc waves in 2″ diameter solid pipe, the modified structure with a TM_{11} attenuation coefficient that is larger by a factor of 100 should tolerate a bending radius of 1,750 feet for the same heat loss increase. (The 50,000 mc attenuation coefficients for ideal 1″ and 2″ copper pipes are 14.8 and 1.79 db/mile respectively.)

For estimating purposes, the ratio of the extra loss per unit line length in the bend region to the straight line loss may be calculated for solid round pipes from the relation

$$\frac{\alpha_B - \alpha_s}{\alpha_s} = \frac{2.5 \times 10^{10} a^6}{\lambda_0^3 R^2} \tag{11}$$

and the absolute increase in attenuation due to a bend is*

$$(\alpha_B - \alpha_s) = \frac{9.7 \times 10^5 a^3}{\lambda_0^{3/2} R^2} \text{ (db/meter for copper guide)} \tag{12}$$

* For small bend radii and bend angles less than θ_c, this relation gives a greater loss than the correct value. See Figs. 22 and 23 of Reference 8 and also see Reference 2.

where α_s = straight line attenuation coefficient
α_B = bend region attenuation coefficient
a = guide radius
R = bending radius
λ_0 = free space wavelength

The approximations used in deriving (11) and (12) are good when the operating frequency is at least 50 per cent greater than cut off for TE_{01}. For guides modified to have higher TM_{11} attenuation both (11) and (12) may be divided by the factor

$$\frac{\overset{m}{\alpha}_{TM_{11}}}{\overset{m}{\alpha}_{TE_{01}}} \cdot \frac{\overset{0}{\alpha}_{TE_{01}}}{\overset{0}{\alpha}_{TM_{11}}} \tag{13}$$

where α^m and α^0 denote attenuation coefficients for the modified guide and solid round guide respectively. On the assumption that the mode coupling is the same in the modified guide as in the solid round guide, use of (13) with (11) or (12) provides an estimate of bend losses in modified circular-electric waveguides.

For a fixed ratio of bend-region attenuation to straightline attenuation, the allowable bending radius varies inversely as the square root of the ratio of TM_{11} heat loss to TE_{01} heat loss, varies inversely as $\lambda_0^{3/2}$, and varies directly as the third power of guide radius.

For fixed bending radius, the *absolute* bend loss varies inversely as $\lambda_0^{3/2}$; since the straight line TE_{01} loss varies directly as $\lambda_0^{3/2}$, bend losses tend to equalize the overall heat loss versus frequency characteristic of the waveguide.

IMPROVED FORMS OF CIRCULAR ELECTRIC WAVEGUIDE

In the preceding discussion it has been indicated that added dissipation for the unused modes of propagation has the effect of decreasing signal losses and of reducing the interference effects associated with mode conversion. Dissipation can be introduced to the unused modes of propagation through the addition of mode filters at intervals along the line, but it appears very desirable to introduce the dissipation to the unused modes on a continuous basis. Several ways of making the line lossy to the non-circular electric modes have been found, and one is illustrated in Figure 26. The copper rings lie in planes transverse to the direction of propagation and provide the conductivity required as a boundary for the circular electric wave family. Successive rings are insulated from each other, however, and the guide provides very poor conductivity in the longitudinal direction. All modes other than the

circular electric wave family have wall currents in the longitudinal direction and experience considerably increased loss in the spaced-ring structure compared to a solid-walled waveguide. A. G. Fox first observed that the spaced ring structure could be used to transmit circular electric waves around bends, and since that time additional work has been carried out by A. P. King and M. Aronoff. The observed loss for the spaced-ring structure under proper conditions was observed to be about 60 per cent more than the theoretical loss for an ideal copper tube, whereas the observed loss for the unused modes of propagation was on the order of 1,000 to 5,000 times the circular electric wave value.

The spaced-ring structure therefore has the electrical properties we seek. The higher-order circular waves exist with losses comparable to their values in a solid copper pipe, but fortunately the magnitudes of conversion between the waves of the circular-electric family have been found to be small. The spaced-ring structure does present some difficult problems with regard to fabrication.

An analogous structure composed of a continuous helical conductor supported within a lossy housing has electrical properties which approximate those of the spaced-ring structure, and the helix should be considerably easier to manufacture in long lengths. The helix might be expected to support a wave-type approximating the circular electric wave both from the standpoint of field distribution and loss when one observes that a helix of very small pitch presents almost circumferential conductivity as required by the circular-electric wave, and the very small longitudinal component necessary due to the finite wire size tends toward zero as the helix pitch tends toward zero. James A. Young of these laboratories has constructed helices in the 2 db/mile waveguide size (4.73″ diameter at 9,000 mc) and found a heat-loss coefficient on the order of 1.75 times the theoretical value for ideal copper pipe. These large ex-

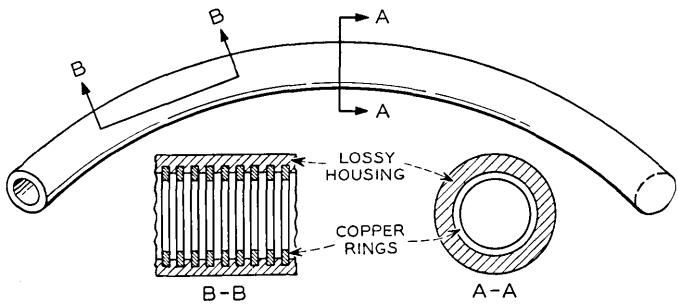

Fig. 26 — Spaced-ring circular electric waveguide.

perimental helices were known to be imperfect, and in a smaller diameter of helix, losses within 15 per cent of the theoretical values for perfect copper pipe were achieved. These results lead us to regard the helical line as a very promising medium for circular electric waves.

SURFACE ROUGHNESS

Since frequencies on the order of 50,000 mc are desirable for low-loss waveguide use, it was recognized that roughness of the surface at the waveguide walls might appreciably increase the heat loss in a practical waveguide. The first approach to this problem was made by W. A. Tyrrell using sections of 5 diameter copper pipe from the experimental line. Tyrrell measured the heat-loss coefficients of the pipe when used as a resonant cavity at 9,000 mc in lengths on the order of 4 to 8 feet. Carefully selected resonant conditions were employed to avoid bringing the unused modes of propagation into resonance at the same time that the circular electric wave was resonated. Whereas Tyrrell observed that the heat loss coefficient in the pipe as originally drawn was about 21 per cent higher than the value computed using the measured dc conductivity, he found that rotary* grinding and polishing the inner surface of the guide reduced the excess loss to about 12 per cent. Tyrrell also observed that commercially drawn brass and 2S aluminum tubing of approximately the same dimensions showed measured losses 11 per cent and 20 per cent respectively greater than the value predicted from the measured dc conductivity. Therefore, the indication from Tyrrell's work was that surface roughness did indeed account for increased losses even at 9,000 mc, and that the excess losses could be reduced either by polishing the surface or through the use of lower conductivities (which have the effect of increasing the skin depth).

A parallel approach to the measurement of surface-roughness effects was made by A. C. Beck and R. W. Dawson, also at 9,000 mc.[9] Beck and Dawson used small wire samples as the center conductor of a coaxial cavity† and found that commercially drawn copper, aluminum and silver wires showed loss values 10 per cent and 15 per cent higher than those expected from the measured dc conductivity. By mechanically polishing

* Because the wall currents for the circular electric wave are circumferential, the longitudinal surface scratches produced by drawing are in the worst possible orientation. Polishing was carried out in a rotary manner so that the current would not cross the scratches so induced.

† In a coaxial, the currents are longitudinal, as are the scratches from drawing, so the measurements in the coaxial would be expected to show somewhat less excess loss due to surface roughness than the measurements made in circular-electric waveguide cavities.

these same wires the losses were reduced to 5 to 8 per cent above the dc values and, by electropolishing, copper wires were brought within 2 per cent of the theoretical value.

An investigation of the effects of aging on the 9,000 mc conductivity of copper surfaces was conducted jointly by Dawson, Tyrrell and Beck, using wires which were measured in the coaxial cavity.[9] They found (1) that the conductivity of untreated surfaces remained essentially stationary when stored indoors; (2) that the conductivity of bare electropolished surfaces deteriorated (over a period of months) to a value comparable to that for commercially drawn wire; and (3) that a tight bonding coating very greatly slows down the aging process.

Recent measurements made in the vicinity of 50,000 mc on a sample of commercially-drawn fine-silver waveguide have indicated that surface roughness effects increased the loss values by approximately 20 per cent.

The overall conclusion would appear to be that it is now feasible commercially to produce waveguide surfaces which have excess losses due to surface roughness no greater than 20 per cent even at frequencies near 50,000 mc, and that by refining the manufacturing method it may be possible to approach the ideal value more closely. In order to avoid aging effects, a tight-bonding coating and a protective atmosphere may be desirable.

CIRCULAR ELECTRIC WAVE AND MILLIMETER WAVE TECHNIQUES

A whole new line of components and techniques are required to carry out in the millimeter wave region the same functions that have always been required in a system application, i.e., power generation, amplification, frequency-band separation, hybrid division, amplitude and phase equalization, and so forth. With a view to ultimate use in a waveguide transmission system, basic research has been done on all of these elements over a period of years at these Laboratories.

The importance of primary oscillators and amplifiers to millimeter-wave systems is self-evident, and work has begun going forward in this difficult field under the direction of J. R. Pierce. The results have been published during the last several years. Pierce made the first 6 mm reflex klystron oscillator.[15] The first 6-mm amplifier, made by John Little,[10] was a travelling wave tube using a conventional helix and it provided 3 db gain. Later, S. Millman[11] introduced the space-harmonic type of travelling-wave amplifier and built several which had over 20 db gain near 6 mm. R. Kompfner[12] devised the backward-wave type of travelling-wave oscillator and built a tube which oscillated from about 5 mm to 8 mm. A. Karp[13] devised a simple structure for backward-wave

oscillators which has already resulted in 5 to 6 mm oscillators and which appears suitable for use at shorter wavelengths.* Recently, E. D. Reed produced a 5 to 6 mm reflex klystron.*

R. S. Ohl has continued his pioneering work on point-contact crystal converters and has made units for use at 6 mm having conversion losses of less than 8 db and output noise ratios on the order of three times basic thermal noise. Ohl also made point contact silicon units for use as harmonic generators to permit the conversion of 24,000 mc power to 48,000 mc power. His harmonic generators have proven invaluable as a source of millimeter wave power — essentially all of the radio research work done to date has been carried out using them.

In order to evaluate crystals and millimeter wave oscillators, it is essential to have an absolute power reference in the millimeter region, and work* has been done by W. M. Sharpless to establish such a reference.

Up to the present time all of the amplifiers, oscillators and other circuit elements have employed dominant-mode rectangular waveguides in order to simplify the circuit design. Therefore, it is of importance to know how to transform a signal from a dominant-mode rectangular guide to the circular electric wave in round pipe. The first circular-electric-wave transducer made in these Laboratories was designed by A. P. King and had the form sketched in Fig. 27. This transducer is of the general type in which the metallic boundary of the waveguide is shaped to force the field lines in the cross section of the guide into the pattern characteristic of the desired output wave. In Fig. 27 the dominant-mode rectangular guide at the left end is gradually tapered to the sector of a circle; the size of this sector is small enough so that only one wave type can exist at this point, and the electric field lines are arcs of a circle. Next, the angle of the sector is gradually increased along the axis of propagation until at one point a cross section of the guide has the shape of a half circle. The size of the sectoral angle is continually increased, however, until finally the metallic sector of the circle disappears as a radial vane. When the taper is done gradually (an overall length of approximately 10 to 15 wavelengths) the electric field lines remain arcs of a circle as

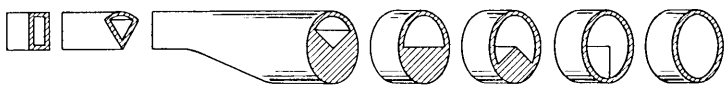

Fig. 27 — Circular-electric wave transducer (due to A. P. King).

* A portion of this work was carried out under Office of Naval Research Contract Nonr 687(00).

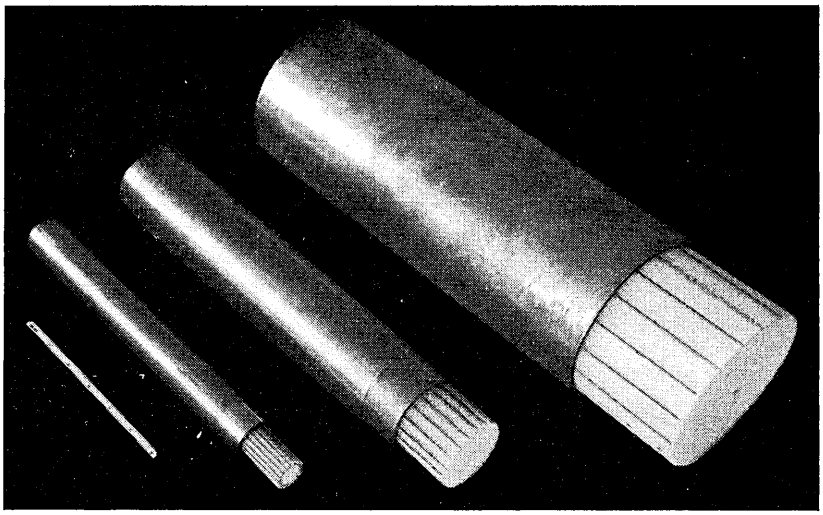

Fig. 28 — Mode filters which pass only circular electric waves.

they were in the sector at the left-hand end of Fig. 27, and the circular-electric wave emerges in the round pipe. This type of transducer has been shown to have transfer losses from dominant-mode rectangular guide to circular-electric wave in round pipe of approximately 0.3 db at 24,000 mc. Similar models have been made by A. G. Fox for use at 48,000 mc.

Another important component is the mode filter previously referred to and which attenuates all wave types other than the circular electric wave family. One type of mode filter to perform this function is the spaced ring structure of Fig. 26 and another type, due to A. P. King, consists of resistive sheets along radial planes as shown in the photograph of Fig. 28. The circular electric wave family has no electric field in a radial direction or in a longitudinal direction. All other wave types, however, have radial-electric or longitudinal-electric field components and experience attenuation due to the presence of the resistive sheets.

The coupled-wave type of transducer sketched in Fig. 29[2] is useful in connecting from dominant rectangular guides to the circular electric or other modes in round guide. This type of transducer makes use of the fact that the various modes in the multimode guide have unequal phase constants. The transfer of power from rectangular guide to the round guide takes place only to the particular round-guide mode whose phase constant is equal to that of the wave in the rectangular guide. This type

of transducer has the same geometric appearance (aside from exact dimensions) for any mode in the round guide and is attractive in that it presents a matched impedance to all the modes of propagation. This property may be used to combine a series of signals onto different modes in a single transmission line. The coupled wave transducer may also be employed to multiplex a series of frequency bands into one pipe. We may wish to employ the long-distance waveguide over the frequency band from perhaps 35,000 mc to 75,000 mc and will require a series of transducers to go from dominant mode guides in various portions of this band to the circular-electric wave in the round guide. Frequency-selective coupled-wave transducers may be employed in the manner sketched in Fig. 30 to multiplex these frequency bands into the pipe for the long distance transmission.

A. G. Fox[14] has shown that dielectric waveguides are attractive as a flexible connecting link for terminal equipment in the millimeter wave region and may also be employed in circuits such as hybrids.

On all of these items of millimeter wave technique and multimode waveguide technique, individual publications will appear as soon as the work has reached the point where this becomes possible.

MODULATION METHODS

The modulation method to be used for the transmission of intelligence on a waveguide system will probably be dominated by the conversion-reconversion phenomenon already discussed. In order to evaluate the

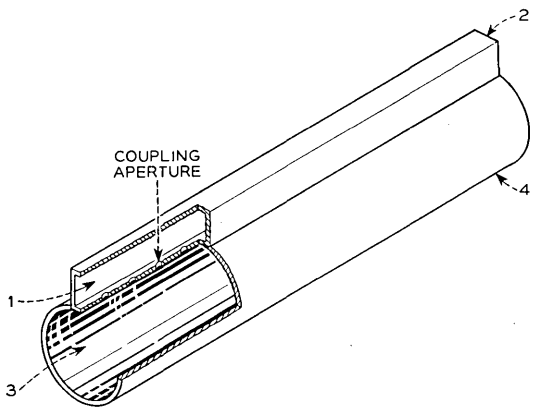

Fig. 29 — Coupled-wave transducer for generating circular-electric or other waveguide modes.

interfering effects of the conversion process, it is important to take into account the time relations between the signal components and the reconverted wave components. In general, it seems clear that reconverted energy returning to the signal mode within a time interval appreciably less than the reciprocal of the highest modulation frequencies will be less interfering than reconverted wave components separated from the signal components by time intervals on the same order as the reciprocal of the modulation frequencies. In practice, a distribution of reconversions will be obtained from a line containing randomly disposed imperfections. As discussed in connection with equation (1), an approximate upper limit can be placed on the time separation between the signal energy and the reconverted wave energy. Table IV records the "upper-limit" time delay for the 5″ diameter experimental 9000 megacycle waveguide

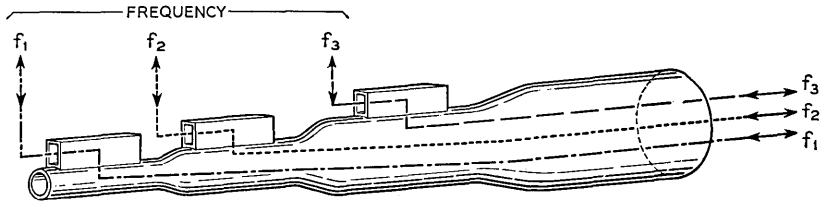

Fig. 30 — Proposed method of multiplexing several frequency bands in the circular electric waveguide.

and for a 2″ diameter pipe used at a frequency of 48,000 mc. The negative time delays represent reconverted-wave components which precede the signal at the receiving end of the system. These calculations suggest that reconverted wave energy might precede the signal by almost ½ microsecond or lag the signal by approximately 1 microsecond in the 5″ diameter 9,000-mc line. Photographs of the observed pulse transmission shown in Fig. 15 indicate negligible reconverted-wave energy at more than one-half these time intervals, and therefore it appears that the method of estimating (described in connection with equation 1) is pessimistic. However, the "upper limit" time delay does provide a convenient way of comparing the experimental line with a proposed 2″ diameter line used at 48,000 mc. Observe in Table IV that the 2″ diameter line will show appreciably less delay for the reconverted wave components than is to be expected from the 5″ diameter line. This is due to higher attenuation in the unused modes and to smaller differences between the group velocities of the various modes in the 2″ diameter pipe. "Upper limit" time delays in the range 1 to 100 millimicroseconds

TABLE IV — "UPPER LIMIT" TIME DELAY FOR RECONVERTED ENERGY COMPONENTS IN SOLID PIPE WITHOUT MODE FILTERS

Mode	"Upper Limit" Time Delay	
	4.732″ Dia. Pipe at 9000 Mc	2.0″ Dia. Pipe at 48,000 Mc
TE_{11}	-0.456×10^{-6} second	-0.0137×10^{-6} second
TE_{21}	−0.097	−0.0035
TE_{31}	+0.0357	+0.0014
TE_{12}	1.093	+0.111

are indicated for the 2″ diameter line and may also be regarded as pessimistic estimates by virtue of our experience with the 5″ diameter line.

It is to be emphasized that the time delays of Table IV apply to solid round guide without mode filters. The addition of mode filters in the solid pipe would increase the loss for the unused modes and appreciably reduce the "upper-limit" time delays for reconverted wave energy. For the improved forms of circular-electric waveguide the increased losses for the unused modes would result in similarly reduced time delays; typical factors of reduction might run between 100 and 1,000 for spaced-ring and helical waveguides compared to the data for solid pipe given in Table IV.

Taking into consideration both conversion-reconversion effects and some equalization of delay-distortion due to waveguide cut-off, it would appear likely that baseband widths on the order of 100 to 1000 mc could be employed in a 2″ diameter line.

The character of the reconverted wave interference will be as discussed in connection with Figs. 15 and 18 if the modulating wave of the signal carrier contains frequencies near the reciprocal of the "upper-limit" time delay for the particular waveguide used.

The magnitude of the reconverted wave energy has been discussed in connection with Fig. 21, and it has been observed that ratios of signal power to reconverted-wave power on the order of 20 db may be expected on a well engineered waveguide line having 60 db of heat loss for the signal wave. Although the time relations were not specifically taken into account in this discussion, it seems likely that the order of magnitude of the reconverted-wave power will be unaltered by more precise analysis.

One might therefore conclude that the modulation method to be used in a waveguide system must be one which will tolerate large amounts of signal interference. Pulse code modulation is one arrangement of signaling which will tolerate such interference, and undoubtedly there are others as well.

Another conclusion the writer has reached is that some form of signal regeneration is likely to be required at each repeater, and that the modulation process should be selected in such a way as to permit this regeneration with limited complexity.

CONCLUSION

Single-mode waveguide is unattractive as a long-distance communication medium due to limited bandwidth and either unreasonable size or excessive loss.

The circular-electric mode in a 2″ diameter round pipe has a theoretical attenuation coefficient of 2 db per mile for carrier frequencies near 50,000 mc. Delay distortion due to waveguide cut-off will require equalization if baseband widths on the order of 500 mc are to be provided on repeater spacings of 25 miles. Using frequency-division multiplex, such a waveguide might be exploited over the 40,000-mc band from 35,000 to 75,000 mc, for which the theoretical attenuation coefficients are 3 db/mile and 1 db/mile respectively.

Transmission experiments were conducted at 9,000 mc in a 500-foot copper pipe 4.73″ I.D. for which the theoretical circular-electric wave loss was 1.9 db/mile. Under favorable conditions the observed losses were within 25 per cent of the theoretical value; under unfavorable conditions (which are not likely to occur in practice) the observed losses were as high as 75 per cent greater than the theoretical value. Surface roughness accounted for losses about 20 per cent above the theoretical value, and the remaining excess losses were due to conversion of energy to other modes of propagation. Direct observation of the power transferred between modes in the 500-foot line confirmed the latter conclusion.

Other observations in the 500-foot line showed that the mode conversion process produced a signal-distortion or signal-crosstalk effect through reconversion of energy from the unused modes of propagation back to the signal (circular electric) mode. This type of interference seriously limits the capabilities of bare copper pipe for use as a long-distance communication medium. However, it has been shown experimentally and theoretically that the effects of the conversion-reconversion process are greatly reduced through the addition of mode filters, which absorb energy present in the unused modes of propagation.

The combination of solid pipe plus mode filters remains attractive as a communication medium. Average losses for the unused modes of propagation should be on the order of 500 times the loss to the signal wave, in order that the conversion-reconversion effects be tolerable in

the presence of conversion losses of the magnitude observed in the 500-foot experimental line.

Improved forms of circular-electric waveguide have been demonstrated. These waveguides have the common property of providing very large attenuation coefficients for the unused modes of propagation while maintaining essentially the same circular-electric wave attenuation as in solid pipe. Thus, these guides are essentially continuous mode filters. Structurally, such a medium may consist of a series of metallic rings supported in a lossy housing (Fig. 26), or may consist of a helix supported in similar manner. The helical circular-electric waveguide appears more attractive from the standpoint of ease of fabrication.

Where bends must be sharp, they may be negotiated in a number of special ways. Gradual bends may be introduced in the improved forms of circular-electric waveguide with modest increase in heat loss. For the 2″ diameter guide at 48,000 mc, it is estimated that a bending radius of about 2,000 feet would double the heat loss of a helical or spaced-ring waveguide. For a 1″ diameter helical or spaced-ring guide (theoretical heat loss of 15 db/mile in solid pipe) the corresponding bending radius is about 200 feet. The extra heat loss due to bending varies inversely as the square of the bending radius.

The type of modulation to be used in a waveguide system will probably be dominated by conversion-reconversion effects. An upper limit on the time delay between the signal component and the associated reconverted-wave components lies in the range 1 to 100 millimicroseconds for 2″ pipe at 48,000 mc without mode filters; the addition of mode filters or use of the helical circular-electric guide should reduce these time delays by factors of 100 or more. Thus, it appears likely that basebands on the order of 500 mc or more should be usable.

It is concluded that a waveguide signalling method must be capable of withstanding large amounts of signal interference, and that some form of regeneration is likely to be required at each repeater.

ACKNOWLEDGMENT

The encouragement of R. Bown, H. T. Friis, and J. C. Schelleng is gratefully acknowledged. The contributions of numerous colleagues (as noted throughout the paper) form the building blocks without which the present understanding of waveguide transmission could not have been obtained.

Appendix

Theoretical Analysis for Continuous Mode Conversion

Whereas the travelling-pulse type of theoretical analysis utilized in the body of this paper can be extended to a realistic spacial distribution of conversion points and to a series of modes instead of only one, J. R. Pierce suggested that the assumption of uniform mode conversion along the axis of propagation would lead to a solution in closed form and would probably show the general properties being sought. This suggestion was adopted and P_1 is designated as the signal power, P_x as the power in the unused mode, and P_n as the power which has transferred from mode-x back to the signal mode, mode 1. We assume quadrature addition of conversion components, and write a series of differential equations expressing the *power flow* between the modes along the axis of propagation, including the heat loss effects:

$$\frac{dP_1}{dz} = -a_{1h}P_1 - a_{1x}P_1 \tag{9}$$

$$\frac{dP_x}{dz} = -a_{xh}P_x - a_{x1}P_x + a_{1x}P_1 + a_{1x}P_n \tag{10}$$

$$\frac{dP_n}{dz} = -a_{1h}P_n - a_{1x}P_n + a_{x1}P_x \tag{11}$$

in which the symbols have the following definitions:

a_{1h} = the heat loss coefficient − mode 1
a_{1x} = the mode conversion coefficient from mode 1 to mode x
a_{xh} = the heat loss coefficient − mode x
a_{x1} = the mode conversion coefficient from mode x to mode 1
z = distance along the axis of propagation

Note that the above heat loss coefficients are those associated with power rather than attenuation coefficients associated with amplitudes, ($2\alpha_1 = a_{1h}$, $2\alpha_x = a_{xh}$). The above equations also imply mode conversion in the forward direction only.

In a phenomenological way, these equations represent the decay of power in the signal mode and the build up of power in both the unused mode x and reconverted energy P_n in mode 1. The general plan is to solve these equations for P_x and P_n in terms of the input wave power. P_n is maintained separate mathematically from P_1, even though both of them are in the same mode, so that we can clearly identify the energy which has been at one time in the unused mode x.

For mathematical solution, (9), (10) and (11) may be put in the following form:

$$\frac{dP_1}{dz} + \alpha P_1 = 0 \tag{12}$$

$$\frac{dP_x}{dz} + \beta P_x - a_{1x}P_1 - a_{1x}P_n = 0 \tag{13}$$

$$\frac{dP_n}{dz} + \alpha P_n - a_{x1}P_x = 0 \tag{14}$$

in which $\alpha = a_{1h} + a_{1x}$

$$\beta = a_{xh} + a_{x1}$$

The general solution for (12), (13) and (14) is given by the following

$$P_1 = k_1 \varepsilon^{-\alpha z} \tag{15}$$

$$P_2 = k_2 \varepsilon^{r_1 z} + k_3 \varepsilon^{r_2 z} \tag{16}$$

$$P_n = -k_1 \epsilon^{-\alpha z} + \frac{k_2}{a_{1x}} (r_1 + \beta)\epsilon^{r_1 z} + \frac{k_3}{a_{1x}} (r_2 + \beta)\epsilon^{r_2 z} \tag{17}$$

where

$$\begin{matrix} r_1 \\ r_2 \end{matrix} = \frac{-(\alpha + \beta) \pm \sqrt{(\alpha - \beta)^2 + 4a_{1x}a_{x1}}}{2} \tag{18}$$

The positive sign is to be associated with r_1 and the negative sign with r_2. For the boundary conditions, at $z = 0$,

$$P_1 = P_0$$

$$P_x = 0,$$

$$P_n = 0,$$

that is to say, the input to the transmission medium being zero in both the x mode and reconverted energy mode, then the solution takes the following form:

$$P_1 = P_0 \epsilon^{-\alpha z} \tag{19}$$

$$P_x = \frac{a_{1x}P_0}{(r_1 - r_2)} \epsilon^{r_1 z} - \frac{a_{1x}P_0}{(r_1 - r_2)} \epsilon^{r_2 z} \tag{20}$$

$$P_n = -P_0\epsilon^{-\alpha z} + \frac{(r_1 + \beta)P_0}{(r_1 - r_2)}\epsilon^{r_1 z} - \frac{(r_2 + \beta)}{(r_1 - r_2)}\epsilon^{r_2 z} \tag{21}$$

It is informative to note that $(r_1 - r_2)$ is always positive and is equal to

$$\sqrt{(\alpha - \beta)^2 + 4a_{1x}a_{x1}}$$

We are usually interested in the ratio of the power in the x-mode to the signal power P_1 and the ratio of the reconverted energy P_n to the signal power P_1. These ratios are given by the following expressions

$$\frac{P_n}{P_1} = \frac{(r_1 + \beta)}{(r_1 - r_2)}\epsilon^{(r_1+\alpha)z} - \frac{(r_2 + \beta)}{(r_1 - r_2)}\epsilon^{(r_2+\alpha)z} - 1 \tag{22}*$$

$$\frac{P_x}{P_1} = \frac{a_{1x}\epsilon^{(r_1+\alpha)z}}{(r_1 - r_2)}\left[1 - \epsilon^{-(r_1-r_2)z}\right] \tag{23}*$$

Thus, we have explicit solutions for the uniform transmission medium containing mode conversions.

In order to make the most general study of these relations, we shall express the mode conversion coefficients in terms of the heat loss coefficient in the signal mode — i.e., as the ratio a_{1x}/a_{1h}. This is natural enough physically, for we are interested in the relative magnitudes of the heat loss and mode conversion effects. It is found that knowledge of the ratios a_{1x}/a_{1h}, a_{x1}/a_{1h} and a_{xh}/a_{1h} enables us to completely determine P_x/P_1 and P_n/P_1 in terms of the distance parameter $\epsilon^{-a_{1h}z}$. The latter is the heat loss in the signal mode, another familiar physical characteristic.

CHARACTERISTIC CONDITIONS IN BARE ROUND WAVEGUIDE

In order to use the theoretical relations derived in the preceding section, we need to know typical values of the parameters. In particular, we need to know the magnitude of typical conversion coefficients a_{1x} and values of the heat loss coefficients a_{1h} and a_{xh} for the modes of interest.

One set of heat loss coefficients which is of immediate interest may be made up from the calculated values for the 5-inch diameter round waveguide used in waveguide experiments at Holmdel. Table V shows the ratio of attenuation coefficients for several modes in this line. The circular electric wave TE_{01} has the lowest attenuation coefficient (absolute value

* When using these relations, it is helpful to note that $(r_1 + r_2) = -(\alpha + \beta)$ at all times. Hence $(r_1 + \alpha) = -(r_2 + \beta)$ and $(r_2 + \alpha) = -(r_1 + \beta)$.

TABLE V — RATIO OF ATTENUATION COEFFICIENTS FOR SEVERAL MODES IN 4.732″ DIAMETER. WAVEGUIDE AT $\lambda_0 = 3.0$ CM

Modes	Attenuation Coefficient Ratio α_x/α_1 or a_{xh}/a_{1h}	Modes	Attenuation Coefficient Ratio α_x/α_1 or a_{xh}/a_{1h}
TE_{12}/TE_{01}	2.45	TE_{31}/TE_{01}	12.5
TE_{02}/TE_{01}	3.82	TE_{41}/TE_{01}	17.
TE_{11}/TE_{01}	4.56	TE_{51}/TE_{01}	21.
TE_{22}/TE_{01}	4.63	TE_{61}/TE_{01}	27.
TE_{13}/TE_{01}	6.61	TE_{71}/TE_{01}	34.
TE_{21}/TE_{01}	8.51	TE_{81}/TE_{01}	44.
TM_{11}/TE_{01}	10.78	TE_{91}/TE_{01}	61.
TE_{03}/TE_{01}	11.3	$TE_{10,1}/TE_{01}$	100.

equal to 1.6 db/mile), and the a_{xh}/a_{1h} ratios for the other modes range from 2.5 to 100 times the TE_{01} value. Table V represents a selection of the various modes which can exist in the 5-inch diameter line, but the number tabulated is *not* an indication of the density of the ratio of attenuation coefficients near a given value. Actually, there are eight modes having a_{xh}/a_{1h} ratios in the range 2.5 to 10, 19 modes in the range 10 to 20, and 15 modes greater than 20.

Experimental work reported elsewhere shows that significant conversion takes place between TE_{01} and the TE_{11}, TE_{21}, TE_{31} and TM_{11} modes. There is some likelihood that conversion to TE_{12} takes place, but the magnitude has not been measured. The experience gained by measurement, therefore, shows that most typical conversions occur between TE_{01} and modes having attenuation ratios a_{xh}/a_{1h} in the range 2.5 to 12.

The absolute magnitudes of the conversion coefficients a_{1x} have in some cases been measured directly, and may also be inferred from measurements of total signal attenuation on the 500 ft. experimental line and separate knowledge of the heat-loss values; the inference is that a_{1x} must fall in the range 0.1 to 1.0 a_{1h} for the particular line studied.

REFERENCES

1. S. E. Miller and A. C. Beck, Low-Loss Waveguide Transmission, Proc. I. R E., **41**, pp. 348–358, March, 1953.
2. S. E. Miller, Coupled-Wave Theory and Waveguide Applications, B. S. T. J., **33**, pp. 661–720, May, 1954.
3. M. Aronoff, Radial Probe Measurements of Mode Conversion in Large Round Waveguide with TE_{01} Mode Excitation, presented orally at the March, 1951, I.R.E. National Convention.

4. M. Jouguet, Effects of the Curvature on the Propagation of Electromagnetic Waves in Guides of Circular Cross Section Cables and Transmission (Paris), **1,**, No. 2, pp. 133–153, July, 1947.
5. M. Jouguet, Wave Propagation in Nearly Circular Waveguides: Transmission-Over-Bends Devices for H_0 Waves, Cables and Trans (Paris) **2,** No. 4, pp. 257–284, Oct., 1948.
6. W. J. Albersheim, Propagation of TE_{01} Waves in Curved Waveguides, B. S. T. J., **28,** pp. 1–32, Jan., 1949.
7. S. O. Rice, unpublished work.
8. S. E. Miller, Notes on Methods of Transmitting the Circular Electric Wave Around Bends, Proc. I.R.E., **40,** pp. 1104–1113, Sept., 1952.
9. A. C. Beck and R. W. Dawson, Conductivity Measurements at Microwave Frequencies, Proc. I.R.E., **38,** pp. 1181–1189, Oct., 1950.
10. J. B. Little, Amplification at 6-mm Wavelength, Bell Labs. Record, Jan., 1951.
11. S. Millman, A Spatial Harmonic Travelling-Wave Amplifier for Six-Millimeters Wavelength, Proc. I.R.E., **39,** pp. 1035–1043, Sept., 1951.
12. R. Kompfner, Backward-Wave Oscillator, Bell Labs. Record, Aug., 1953. R. Kompfner and N. T. Williams, Backward Wave Tube, Proc. I.R.E., **41,** pp. 1602, Nov., 1953.
13. A. Karp, Paper presented orally at the June, 1951, Conference on Electron Tube Research.
14. A. G. Fox, New Guided-Wave Techniques for the Millimeter-wave Range, presented orally at the March, 1952, I.R.E. National Convention.
15. J. R. Pierce and W. G. Shepherd, Reflex Oscillators, B. S. T. J., **26,** pp. 460–681, July, 1947.
16. G. D. Sims, The Influence of Bends and Ellipticity on the Attenuation and Propagation Characteristics of the H_{01} Circular Waveguide Mode, Proc. Institution of Electrical Engineers (London), **100,** Part IV, No. 5, pp. 25–34 Oct., 1953
17. S. P. Morgan, Jr., Mode Conversion Losses in Transmission of Circular Electric Waves through Slightly Non-Cylindrical Guides, J. Appl. Phys., **21,** pp. 329–339, April, 1950.

A Governor for Telephone Dials

Principles of Design

By W. PFERD

(Manuscript received July 29, 1954)

This is a report on the development of a new type of governor for regulating the speed of rotary dials. The paper includes derivation of the equations of motion which determine the theoretical speed of the governor during dial run-down and an analysis of the operating characteristics of the governor as influenced by varying friction and input torque. Experimental verification of the relations is presented. A theoretical analysis which explains "governor chatter" or positional instability for friction-centrifugal governors is also given.

INTRODUCTION

Machine switching telephone systems depend on the telephone dial for originating information used in completing a call. During run-down, the dial originates current pulses which operate step-by-step switching equipment or are registered for use in common-control panel or crossbar systems. For reliable functioning of dial pulse controlled switching equipment, the pulses must be closely controlled in frequency and form. Since the pulses are produced during run-down of the dial after release by the customer, the run-down speed most be constant. Friction-centrifugal governors are commonly used to provide this required control of speed.

If the pulses reaching the central office were exactly like those generated by the dial, the designers of dials and central office switching apparatus and circuits would find themselves far less restricted. Unfortunately the dial pulses are distorted by the electrical characteristics of the customer's loop. To compensate for this distortion and insure accurate registration of the pulses at the central office, the dial and central office equipment must be designed to operate to close limits of performance. The designs must also be such that there is negligible change of

adjustment resulting from use, time or weather conditions, which might affect the ability to accurately send or receive dial pulses.

To achieve the accuracy of timing required of dial governors, it was recognized that a general theoretical analysis which defined speed of governors would be beneficial. The late C. R. Moore investigated this problem in the thirties, and derived from theoretical considerations, general equations of motion relating to governors. The relationships derived by the Moore analysis are extremely useful in that they can be used to indicate the influence of various design factors on the performance of a governor. This theory was applied in developing the new governor used in the 7-type dial of the 500-type telephone set* and will be presented in this paper. To better demonstrate the operating characteristics of the new governor, it will be compared with a previous governor which was used in an older type dial. Photographs of the new governor as it is assembled in a dial are shown in Figs. 1A and 1B.

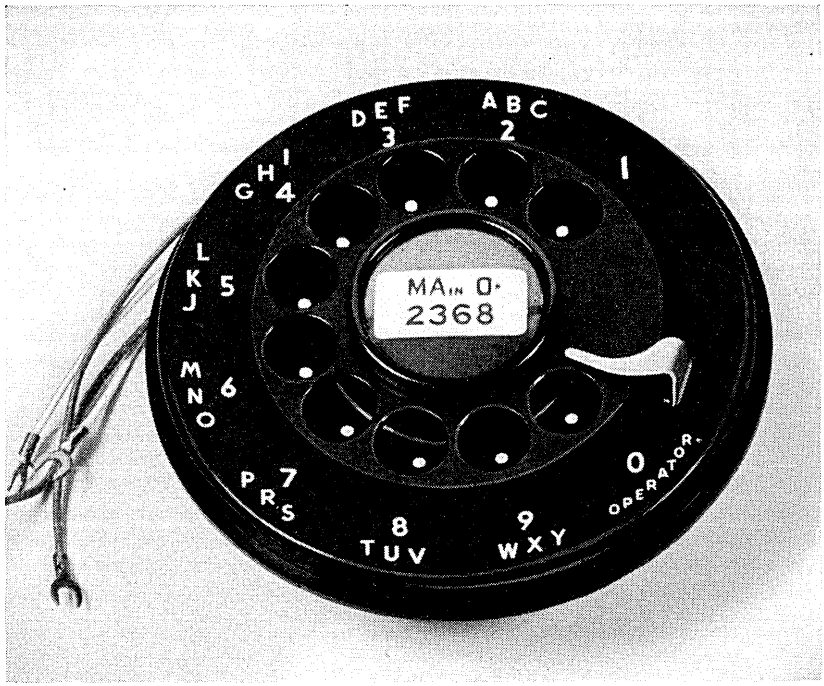

FIG. 1A — Front view of 7-type dial.

* Inglis, A. H., and Tuffnell, W. L., An Improved Telephone Set, B.S.T.J., **30**, pp. 239–270, April, 1951.

DIAL AND GOVERNOR OPERATION

In dialing, the fingerwheel is rotated through an angle proportional to the number being dialed and then released. Energy stored in the motor spring, Fig. 2, causes the mechanism to return to the start position. For each 30° rotation of the fingerwheel during run-down, the intermediate gear rotates one-half revolution and the cam pinion and pulsing cam rotate one full revolution. Once the pulsing pawl is in position, each revolution of the pulsing cam in the run-down direction results in a pulse being placed on the telephone loop. The intermediate gear also meshes with a governor pinion which is coupled to the governor shaft and governor through a spring clutch.* This clutch decouples the governor from the fingerwheel on windup to reduce the windup torque. On run-down the governor rotates two full revolutions for each 30° rotation of the fingerwheel.

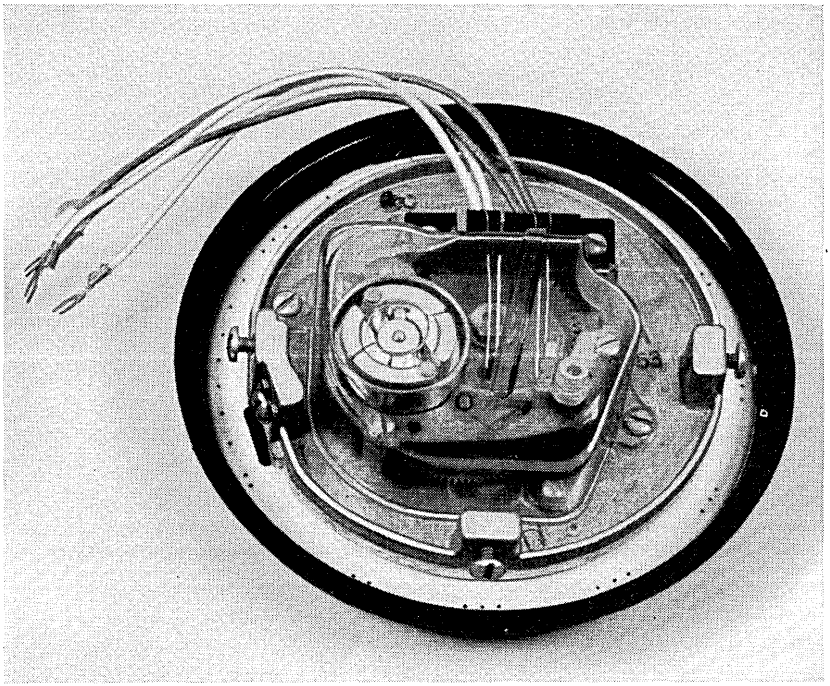

FIG. 1B — Rear view of 7-type dial. Left, speed governor, center, off-normal contact and right, pulsing mechanism.

* Wiebusch, C. F., The Spring Clutch, J. Appl. Mech., Sept., 1939.

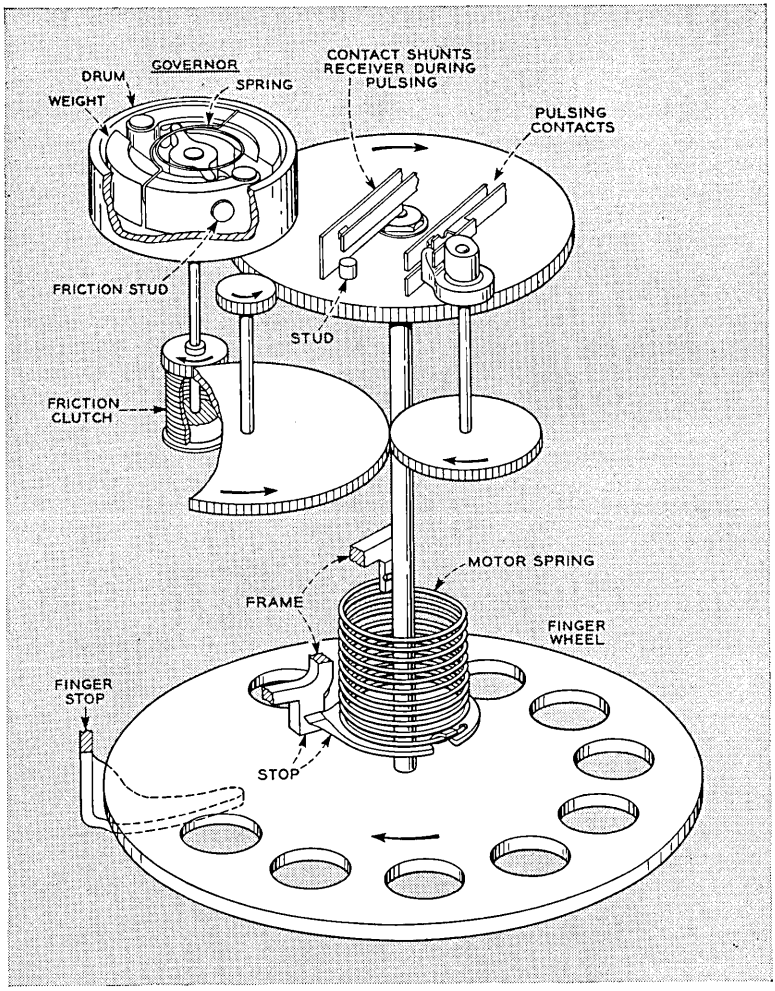

FIG. 2 — Simplified diagram of 7-type dial mechanism; governor appears top left.

As shown in Fig. 3, the weights of the new governor are free to pivot at the ends of the fly-bar which, in turn, is allowed to rotate with respect to the shaft. Rotation is imparted to the system by the drive-bar which presses against each weight at a specific point. As the mechanism begins to rotate during dial run-down, the two weights are caused by centrifugal force to move outward against the tension of a spring. Movement of the

weights about their pivots continues until the friction studs touch the case. At this instant governing begins, and controls the dial speed until the end of run-down. The speed attained by the governor will be dependent on the friction between the studs and case, the magnitude of the driving torque, and the tension to which the spring of the governor is adjusted.

In the schematic of the new governor, Fig. 4, the driving force is designated as F. By applying this driving force between the weight pivot and the center of the governor, the mechanism behaves during operation as a true friction-centrifugal governor and also as a brake. This configuration results in a gain in the ability of the governor to resist the increase in speed which normally results from an increase in the applied rotational force. The drive-bar force, F, and the torque produced by the stud-to-case force about the pivot assists the centrifugal force in pressing the friction studs against the case.

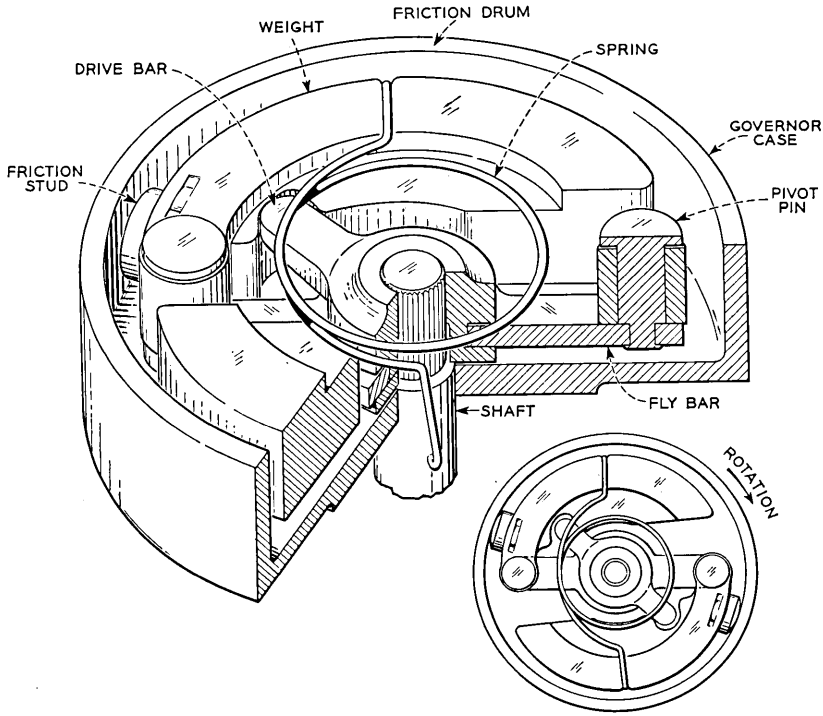

FIG. 3 — The 7-type dial drive-bar governor.

The comparative design shown in Fig. 5 is a conventional fly-bar type governor consisting of two weights each pivoted at the end of a fly-bar which is fixed to a shaft. As the shaft accelerates during run-down, the weights move outward under the influence of centrifugal force and are restrained in their motion by the tension of the governor spring. At a certain speed the friction studs contact the inner surface of the governor case. The governor gradually decelerates until the input torque to the governor is balanced by the stud-to-case frictional loss and the governor shaft and dial theoretically rotate at constant speed. It will be noted in this configuration that only the torque produced by the stud-to-case

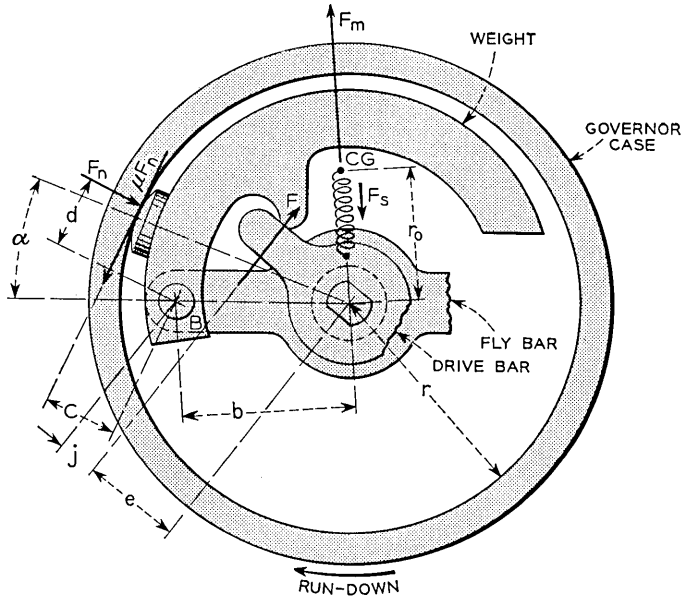

Fig. 4 — Schematic of drive-bar governor.

F — Force applied by torque on governor weights
F_n — Normal force of case acting on studs
F_s — Force exerted by spring when studs touch case
F_m — Centrifugal force acting at center of gravity of weight
μ — Coefficient of friction
I_0 — Moment of inertia of the governor about center shaft
ω — Angular velocity of governor
ω_0 — Critical angular velocity at which studs just touch the case
m — Mass of each weight
r_0 — Radius to the center of gravity of each weight
r — Radius of governor case
α — Stud angle
Neg. Rotation — Run down of governor

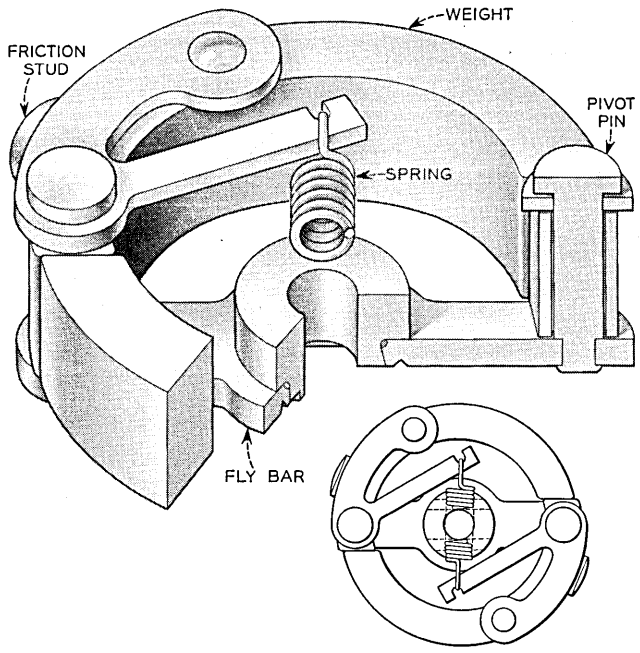

Fig. 5 — Fly-bar type governor.

friction about the pivots aids the centrifugal force in pressing the studs against the case.

EQUATIONS OF MOTION

In deriving the general equations of motion for a governor, two assumptions are made concerning the action. During the interval of time that the governor is approaching the critical velocity when contact of the friction surfaces first occurs, the motion is assumed to be that of a simple fly-wheel, constantly accelerating. The angular velocity of the governor during the time from rest to the critical velocity, ω_0, is then given by

$$\omega = \frac{Gt'}{I} \tag{1}$$

where G = applied torque
t' = time from start of motion
I = moment of inertia of the governor assembly about the shaft

After stud-to-case contact occurs, it is assumed that there is no further pivoting of the weights or fly-bar, and the assembly rotates as a rigid body. During this time the general equation for angular velocity is

$$\omega = q \tanh\left(\frac{ht}{q} + ln\sqrt{A}\right) \tag{2}$$

where q = regulated or final angular velocity
h = design constant
A = design and adjustment constant
t = time measured from the moment of initial braking

The derivation of this general speed equation as it applies to the new 7-type dial governor is given in Appendix I. For this drive-bar type governor the following terms apply:

$$h = \frac{G(d - \mu c) + M\mu\omega_0^2 - \mu rjG/e}{I_0(d - \mu c)} \tag{3}$$

$$g = \frac{M\mu}{I_0(d - \mu c)} \tag{4}$$

$$A = \frac{q + \omega_0}{q - \omega_0} \tag{5}$$

$$q = \sqrt{\frac{h}{g}} = \sqrt{\frac{G(d - \mu c) + M\mu\omega_0^2 - \mu rjG/e}{M\mu}} \tag{6}$$

The derivation of the theoretical equation for speeds in excess of the critical speed for the comparative fly-bar design results in the following relationships:

$$\omega = q \tanh\left(\frac{ht}{q} + ln\sqrt{A}\right)$$

where

$$h = \frac{G(d - \mu c) + M\mu\omega_0^2}{I_0(d - \mu c)} \tag{7}$$

$$g = \frac{M\mu}{I_0(d - \mu c)} \tag{8}$$

$$q = \sqrt{\frac{h}{g}} = \sqrt{\frac{G(d - \mu c) + M\mu\omega_0^2}{M\mu}} \tag{9}$$

The form of the general speed equation is the same for all types of

friction-centrifugal governors but each particular governor will have different terms in the values for g, h and q. The theoretical equation of motion can be used to calculate the speed of the dial at any time, t, after the critical velocity is reached or the time required to reach any given speed once the governor studs touch the case. The equation shows that for large values of t, ω approaches q, so that steady state speed is given by the value of q for each type governor, and is in terms of the operating values and design constants of the mechanism.

For the drive-bar governor the steady state speed equation is

$$\omega = q = \sqrt{\frac{G(d - \mu c) + M\mu\omega_0^2 - \mu rjG/e}{M\mu}} \tag{10}$$

THEORETICAL SPEED-TIME CURVES

Drive-Bar Governor

The design constants and physical data given in Table I apply to the drive-bar governor, and were used to calculate the theoretical speed

TABLE I — REFER TO FIGS. 3 AND 4

d = 0.390 cm	I_0 = 7.40 gm cm² (experimental)
c = 0.361 cm	G = 7,500 dyne-cm (steady-state)*
r = 1.180 cm	13,500 dyne-cm (initial)
r_0 = 0.635 cm	μ = 0.25 (assumed)
b = 0.920 cm	m = 3.9 gms
e = 0.498 cm	$K = r_0/r$ = 0.538
j = 0.236 cm	$M = 2mr^2bk$ = 5.38

* Appendix II — Governor Input Torque.

versus time curve shown in Fig. 6. The dial governor is initially adjusted so that signaling is at the rate of 10.0 pulses per second, requiring a steady state governor shaft velocity of 125.6 radians per second. This steady state velocity was used to determine the critical velocity ω_0 by substituting the values in Table I in equation (10):

$$\omega = q = \sqrt{\frac{G(d - \mu c) + M\mu\omega_0^2 - \mu rjG/e}{M\mu}}$$

$$\omega_0 = 121.8 \text{ radians/second}$$

and from equations (3), (4), (5), and (6)

$$g = 0.606 \qquad h = 9{,}520 \qquad h/q = 75.9 \qquad A = 72.7.$$

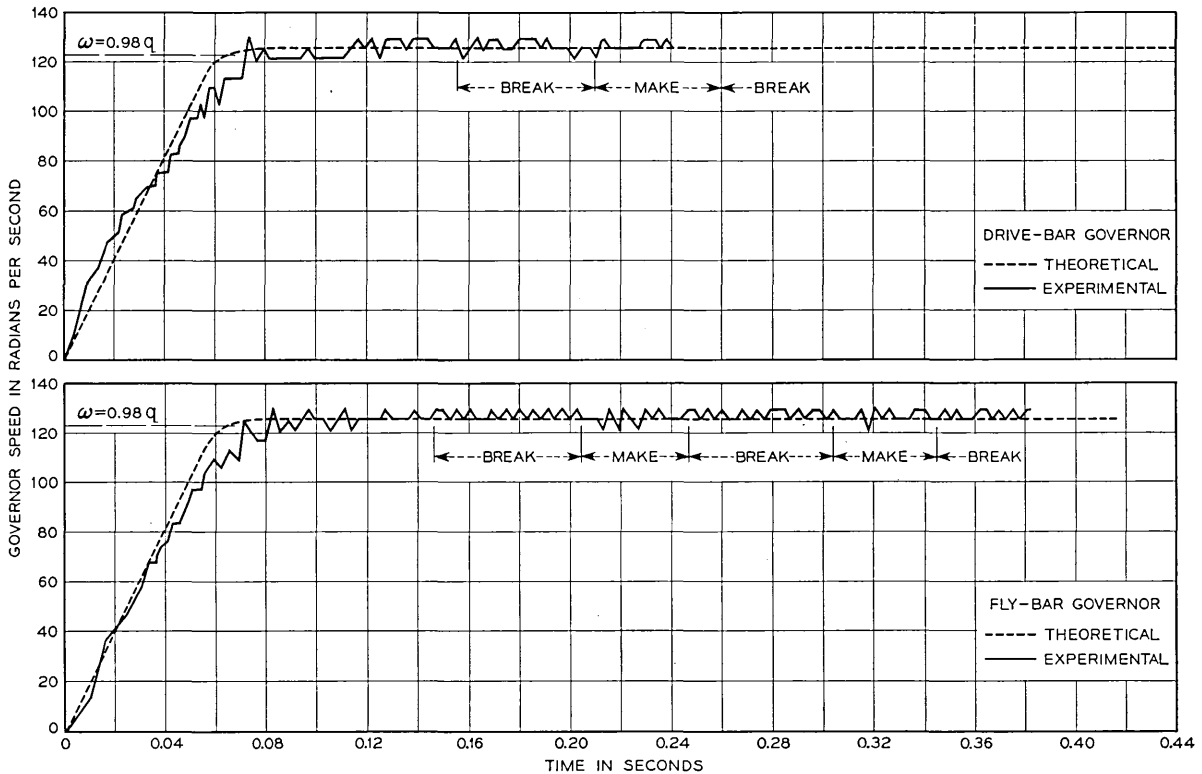

Fig. 6 — Speed curves for the drive-bar and fly-bar type governors.

TABLE II

t	$75.9t + 2.142$	$\omega = 125.6 \tanh (75.9t + 2.142)$
0.000	2.142	121.8
0.004	2.446	123.7
0.008	2.750	124.6
0.012	3.054	125.1
0.016	3.358	125.3
0.020	3.662	125.5

Substituting these values in the general speed equation,

$$\omega = q \tanh\left(\frac{ht}{q} + ln\sqrt{A}\right)$$

gives the velocity of the drive bar governor at any time t measured from the moment the governor reaches the critical velocity, i.e., when the friction studs first touch the inside of the governor case. (Table II.)

For governor speeds from start of rotation up to the critical velocity, it is assumed that the system rotates as a simple fly-wheel, therefore from equation (1)

$$t' = \frac{\omega_0 I_0}{G_I} = \frac{121.8(7.4)}{13{,}500} = 0.0668 \text{ seconds}$$

This time of 0.0668 seconds determines the slope of the straight line portion of the theoretical speed-time curve shown on Fig. 6.

Fly-Bar Governor

The data given in Table III applies to the fly-bar governor shown schematically in Fig. 7. Substituting the values given in this table in the steady state speed equation for the fly-bar governor,

$$\omega = q = \sqrt{\frac{G(d - \mu c) + M\mu\omega_0^2}{M\mu}}$$

$$\omega_0 = 118.5 \text{ radians/sec.}$$

and from equations (7), (8) and (9):

$$g = 0.609 \qquad h = 9{,}560 \qquad h/q = 76.4 \qquad A = 36.35$$

Substituting these values in the general speed equation gives the velocity of the fly-bar governor at any time (t) measured from the moment braking begins, Table IV. For this particular fly-bar design, the

time required to reach the critical velocity is, from equation (1),

$$t' = \frac{\omega_0 I_0}{G_I} = \frac{118.5(7.36)}{13{,}500} = 0.0646 \text{ seconds}$$

EXPERIMENTAL SPEED-TIME CURVES — NORMAL TORQUE

Experimental velocity versus time curves were obtained for the fly-bar and drive-bar governors constructed to the specifications listed in Tables I and III. Data used in determining the true speed versus time picture for the experimental governors was taken from photographic traces exposed on a recording oscillograph. A thin disc having 36 radial slots spaced every 10° was fastened to the end of the governor shaft. Light detected through the slots of the moving disc by the element of a photo tube was used to deflect one of the strings of the oscillograph. The trace

TABLE III — REFER TO FIGS. 5 AND 7

d = 0.390 cm	G = 7,500 (steady state)
c = 0.361 cm	13,500 (initial)
r = 1.18 cm	μ = 0.25 (assumed)
r_0 = 0.635 cm	m = 3.9 gm
b = 0.920 cm	$K = r_0/r$ = 0.538
I_0 = 7.36 gm cm²	$M = 2mr^2bK$ = 5.38

of this string appeared on the photographic paper as a distorted sine wave. The distance between two successive wave peaks represented 10° of rotation of the governor. By noting on the trace the time between peaks, it was possible to determine the average velocity of the governor at 10° intervals after release of the finger-wheel, or start of rotation of the governor mechanism. The experimental speed curves for the drive-bar and fly-bar governors are plotted on Fig. 6 along with the theoretical speed curves.

It was assumed in the theoretical analysis that while accelerating up to the critical velocity, the governor assembly rotates as a simple fly-wheel. This requires that the velocity increase linearly. The theoretical and experimental velocity curves for both type governors during the initial accelerating period show the fly-wheel assumption to be justified. The slope of the velocity curve, or rate of acceleration is generally constant.

For that portion of the theoretical and experimental curves which show speeds from the critical velocity to 98 per cent of rated speed, agreement is not too clearly defined. The theoretical curve is naturally smooth in shape. An oscillating type characteristic appears in the experimental

speed curves of both types of governors. This probably results from the shock and grabbing when the friction studs first touch the case and continues until the forces tending to move the weights outward increase to a value sufficient to hold them against the case for governing.

That part of the experimental curve, during which governing actually occurs at rated speed is relatively smooth through full run-down. Both the fly-bar and new drive-bar governors exhibit excellent speed regulation. The waves present on the trace do not necessarily indicate hunting or vibration since the variations in speed which appear are actually smaller in magnitude than the degree of accuracy present in measuring the experimental photographic trace.

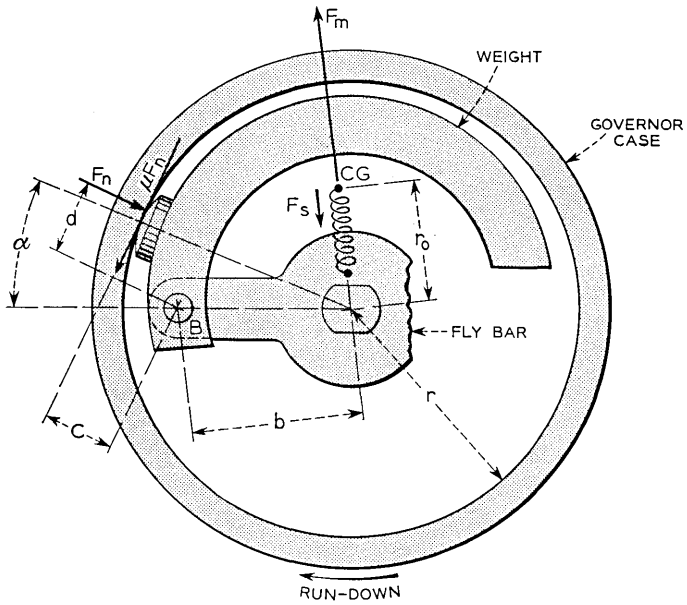

Fig. 7 — Schematic of fly-bar governor.

F_n — Normal force of case acting on studs
F_s — Force exerted by spring when studs touch case
F_m — Centrifugal force acting at center of gravity of each weight
μ — Coefficient of friction
I_0 — Moment of enertia of the governor about center shaft
ω — Angular velocity of governor
ω_0 — Critical angular velocity at which studs just touch the case
m — Mass of each weight
r_0 — Radius to the center of gravity of each weight
r — Radius of governor case
α — Stud angle
Neg. Rotation — Run down of governor

THEORETICAL OPERATING CHARACTERISTICS

The solution of the equation of motion for any governor is based on specific design constants and certain assumed and estimated values. Such dimensions as the governor case inside diameter r, and distance from the shaft center to the weight pivot b, are two examples of design constants. These constants establish the arrangement of the various component parts of the mechanisms and are, therefore, subject to practical manufacturing and space considerations as well as considerations from a speed regulation standpoint. The design constants are in effect static considerations. They are not subject to appreciable variation once established, and on any particular governor do not vary significantly over the life of the governor.

TABLE IV

t	$76.4t + 1.798$	$\omega = 125.6 \tanh (76.4t + 1.798)$
0.000	1.798	118.5
0.004	2.103	121.9
0.008	2.408	123.6
0.012	2.714	124.5
0.016	3.010	125.0
0.020	3.325	125.3
0.024	3.631	125.4

During actual operation there are two factors which directly affect the degree of speed control afforded by a governor; i.e., the coefficient of friction which exists between the governor case and friction studs, and the value of input torque to the governor shaft. Design control over these factors is present, but to a lesser degree than for the design constants mentioned previously. These factors are considered fixed in arriving at a given design but actually vary from governor to governor and over the life of a governor. It is therefore necessary to consider carefully the effect of variations in friction and torque if close regulation of speed is required.

The input torque to the governor shaft will vary because of the dimensional variations of the motor springs, the tolerance permitted for the driving torque at full windup of the dial, dial friction and the variation in pulsing spring forces. These variables appear at the governor as a range of input torques during run-down of the mechanism. For the motor spring used in the 7-type dial, input torque referred to the governor shaft decreases during run-down on the average from 20,000 dyne-cm to 13,000 dyne-cm. Torque required to overcome bearing and gear fric-

tion and the loads imposed by the pulsing mechanism result in an average torque of 7,500 dyne-cm at the governor. Over the life of a dial this input torque at the governor will vary as the dial efficiency varies. Initially the dial mechanism is lubricated and the bearings and gears turn freely. With time and continued operation, the accumulation of dirt and wear products affect the dial so that more torque is needed to drive the moving parts. This causes a decrease in the remainder torque going to the governor.

Another aspect of torque requiring consideration is that resulting from forcing of the finger-wheel during run-down. This action can produce torque values at the governor of the order of 110,000 dyne-cm or a torque of approximately fifteen times that which appears at the governor during normal operation.

The second factor, which can vary during dial life, is the value of the stud-to-case coefficient of friction, μ. Both the drive-bar and fly-bar governors have studs of Ebonite compounded with 40 per cent by weight of hard rubber dust and cases of ASTM B16 brass. Actual service tests show the satisfactory wearing ability of these materials.

Each governor is initially adjusted for speed by changing the tension of the governor spring. At the time this adjustment is made, a particular friction condition exists between the governor studs and case. With time or continued operation there is always the possibility of a change occurring in this friction value. Such factors as very high humidity, lubrication products traveling to the stud operating region, or the accumulation of foreign-material or wear particles may produce different values of friction and hence result in variation in governor and dial speed from the initially adjusted value.

The range of coefficient of friction values μ expected for rubber on brass is from 0.05 to 0.35. These are the extreme conditions produced by oil in the governor case for the 0.05 value and very low unit pressure on a scored brass surface for the 0.35 value. For this study a representative figure for the average stud-to-case friction value was taken to be 0.25.

The problem of variation in steady state governor speed with changes in the coefficient of friction and input torque can be analyzed by considering the derivatives of speed with respect to these values. This is done by operating on the equations for terminal speed.

Speed with Respect to Coefficient of Friction

For the drive-bar governor, the partial derivative of speed, with respect to coefficient of friction, is as follows:

$$\omega = q = \sqrt{\frac{G(d - \mu c) + M\mu\omega_0^2 - \mu rjG/e}{M\mu}}$$

Taking the derivative of ω with respect to μ gives

$$\frac{\partial\omega}{\partial\mu} = -\frac{Gd}{2\omega M\mu^2} \tag{11}$$

where $M = 2mrr_0b$. For the fly-bar governor, $\partial\omega/\partial\mu$, is as follows from

$$\begin{aligned} \omega = q &= \sqrt{\frac{G(d - \mu c) + M\mu\omega_0^2}{M\mu}} \\ \frac{\partial\omega}{\partial\mu} &= -\frac{Gd}{2\omega M\mu^2} \end{aligned} \tag{12}$$

Since the equations are identical for both governors the same considerations exist in holding to a minimum the change in speed caused by a change in the coefficient of friction.

For optimum speed regulation, the partial derivative, $\partial\omega/\partial\mu$, should be a minimum. Small values of $\partial\omega/\partial\mu$ can be obtained by operating on the design constants, controlling the value of μ, or having high governor speeds. Specifically, the design constants m, r_0, r and b should all be large. Inspection of the drive-bar governor schematic, Fig. 4, shows that there are physical limitations to the arbitrary enlargement of these values. Space, manufacturability, and cost of materials must be considered in establishing these terms. In selecting materials for fixing the coefficient of friction value the wearing quality of the materials to be used must be of first consideration. A high governor speed is advantageous but must be weighed against the primary disadvantage of high inertia loads for the entire mechanism.

The terms G and d in the numerator of equations (11) and (12) indicate that low input torque and a small value for d would be desirable For G, one must consider the anticipated change in dial efficiency and variation in torque produced by the pulsing mechanism plus the torque necessary for the governor to maintain adequate speed control.

As shown in the drive-bar governor schematic, Fig. 4, d is the distance from the weight pivot to the normal of the point of contact between the rubber stud and governor case. Its magnitude is controlled by the stud angle α, and the distance between the stud and case when the weights are in the closed position.

As indicated, d should be as small as possible for best regulation with friction change. This requirement imposes a difficult design problem because as d decreases, the stud and bearing hole in the weight approach

each other. A minimum d is therefore fixed by interference of the parts themselves. Inspection of the governor mechanism also shows that even if the interference problem were not limiting, the weight turning angle imposes a further restriction on the d value. The life of a governor is considered to end when the friction material is worn to the point of allowing the weight to touch the inside of the case. Because of this initially large weight motion, little material is provided for wearing and hence the possible life of the mechanism is reduced. In the design of the drive-bar governor, the stud angle, α, was made as small as manufacturing techniques would permit. The stud angle is 22° from the weight pivot with a corresponding $d = 0.390$ cm.

Since both governors under study were designed to operate in the case of the same dial the terms shown in Tables I and III which apply in the friction equations (11 and 12) are identical. This identity of terms indicates that both type governors should exhibit identical frictional characteristics. Fig. 8 represents a plot of the $\partial\omega/\partial\mu$ for various governor input torque values. One set of curves applies. The range of torque values covered by the curves is from zero to the forcing condition at fifteen times normal motor spring torque. Curves for coefficient of friction values, 0.05, 0.10, 0.20, 0.35, are shown as encompassing the extremes in operating range.

Speed with Respect to Input Torque

For the drive-bar governor, the partial derivative of speed with respect to torque is determined as follows from the steady state speed equation.

$$\omega = q = \sqrt{\frac{G(d - \mu c) + M\mu\omega_0^2 - \mu rjG/e}{M\mu}}$$

Taking the derivative of ω with respect to G gives:

$$\frac{\partial\omega}{\partial G} = \frac{1}{2}\left[\frac{G(d - \mu c) + M\mu\omega_0^2 - \mu rjG/e}{M\mu}\right]^{-1/2} \frac{(d - \mu c - \mu rjG/e)}{M\mu}$$

and

$$\frac{\partial\omega}{\partial G} = \frac{1}{2}\frac{(d - \mu c - \mu rj/e)}{M\mu\omega} \tag{13}$$

where

$$M = 2\, mrr_0 b$$

For the fly-bar governor the derivative of ω with respect to G can be

developed similarly to give

$$\frac{\partial \omega}{\partial G} = \frac{1}{2} \frac{(d - \mu c)}{M \mu \omega} \tag{14}$$

Here also the design constants, m, r_0, r and b, the coefficient of friction μ, and the governor speed ω, should all be high values. Minimum change in dial speed would then occur for a given change in the input torque.

Inspection of the equations for $\partial\omega/\partial G$, indicates that it is possible to have perfect torque regulation for at least one value of μ. For this limiting condition, $\partial\omega/\partial G$, would equal zero and equation (13), for the drive-bar governor, equates to

$$d - \mu c - \mu r j/e = 0 \tag{15}$$

and equation (14) for the fly-bar type governor

$$d - \mu c = 0 \tag{16}$$

If these equations were satisfied, there would be zero change in speed for a given change in input torque to the governor. As stated previously μ is predetermined and has in this design a maximum known value of 0.35. A margin of safety is considered by taking $\mu = 0.425$ for the limiting case and equation (15) becomes for the new drive-bar governor

$$d - 0.425\,(c + rj/e) = 0 \tag{17}$$

and equation (16) for the fly-bar governor

$$d - 0.425\,c = 0 \tag{18}$$

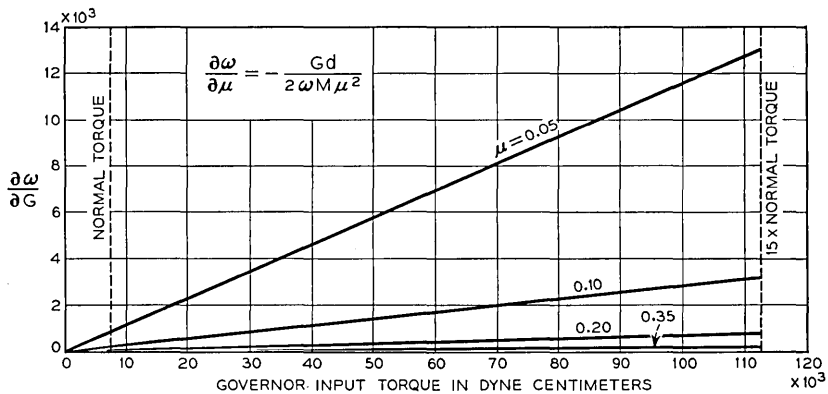

Fig. 8 — Derivative of governor speed with respect to coefficient of friction versus input torque to the governor.

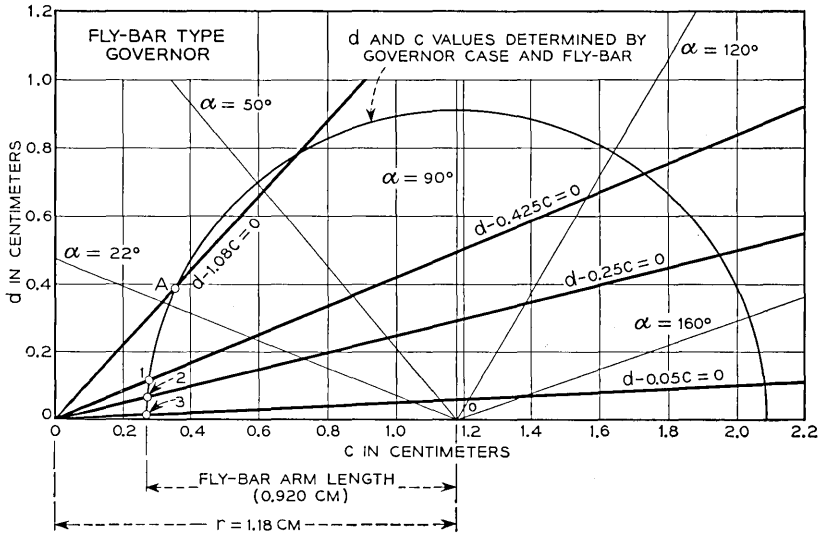

Fig. 9 — Design diagram for fly-bar type governor.

Comparison of these two equations shows a very important difference in the term multiplied by $\mu = 0.425$. For the drive-bar governor, there are four variables which can be operated on to satisfy the equation; i.e., c, r, j and e. For the fly-bar governor only c is available. The importance of these additional terms can be realized when one considers that the value for c results from our choice of d in making $\partial\omega/\partial\mu$ a minimum. For both type governors c is equal to 0.361 cm. Substitution of the d and c values in the limiting equation, (18), for the fly-bar governor does not lead to a solution when $\mu = 0.425$. Solving for this limiting μ in the fly-bar governor gives

$$(0.390) - (0.361)\,\mu = 0$$

$$\mu = 1.08$$

This value of μ is far beyond that encountered in actual governor operation and, in effect, represents useless margin. This is graphically shown in Fig. 9 where all d and c values which conform to the geometry of the fly-bar governor mechanism are plotted as a design diagram. The three straight lines radiating from the origin represent plots of the limiting equation for $\mu = 0.425$, 0.25 and 0.05. The intersection of these lines with the d and c semicircle, noted at points 1, 2 and 3, give the particular angle at which the stud should be located for optimum regula-

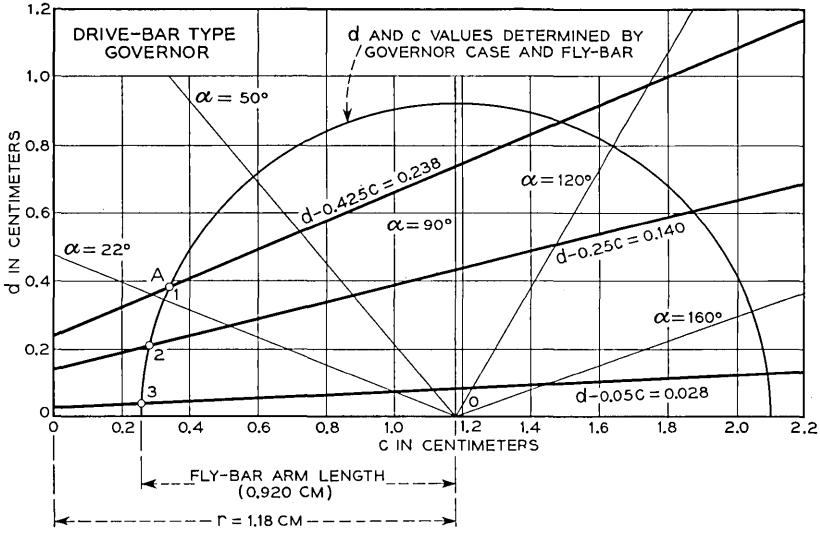

Fig. 10 — Design diagram for drive-bar type governor.

tion for these particular friction values. The point A on the d and c circle denotes the d and c values which result from the stud being at $\alpha = 22°$. The off-set appears because stud-to-case contact is not made on the center line of the stud. The points 1, 2 and 3 are all below the point, A, established by the minimum permissible angle, $\alpha = 22°$. This indicates that the fly-bar governor has its $\partial\omega/\partial G$ equal to zero at some coefficient of friction value higher than $\mu = 0.425$. As previously determined in this value is $\mu = 1.08$.

Fig. 10 represents the design diagram for the new drive-bar governor. Here again, the large semi-circle is a plot of all d and c values which conform to the geometry of this governor mechanism. Point A is determined by the stud angle $\alpha = 22°$. The straight lines represent the limiting equations for $\mu = 0.425$, 0.25 and 0.05 and are shown intersecting the d and c circle at 1, 2 and 3 respectively. For this particular governor, the line representing $\mu = 0.425$,

$$d - 0.425\ (c + rj/e) = 0$$

intersects the d and c circle at the point A. This is possible by making the term $(\mu rj/e)$ equal to 0.238. The intersection of this curve at A indicates that there will be zero change in speed for a given change in input torque to the governor when the stud-to-case coefficient of friction value is 0.425.

The additional terms, r, j and e in the limiting equation make it possible to design the governor for optimum regulation for any particular value of μ desired. Since the term r, the case inside radius, is controlled by space requirements, the terms j and e assume added importance. They are determined by the point at which the drive-bar arms act against the weights and can be made any value required to meet the design objective.

Figure 11 is a plot of $\partial\omega/\partial G$ at various values of μ for the drive-bar and fly-bar governors specified in Tables I and III. It indicates that for any coefficient of friction the new drive-bar governor should exhibit less change for a change in input torque than is possible with the fly-bar governor.

EXPERIMENTAL DATA

To substantiate the theoretical conclusions drawn from the analysis of the steady state speed equations, experimental data were compiled from a number of models of each type of governor. Drive-bar and fly-bar governors, made to the specifications listed in Tables I and III were investigated to determine their response to changes in input torque and changes in the stud-to-case coefficient of friction value, μ. The tests were conducted on 7-type dials manufactured by the Western Electric Company as standard product.

Dial Speed Versus Coefficient of Friction

The theoretical analysis of the equation

$$\frac{\partial\omega}{\partial\mu} = -\frac{Gd}{2\omega M\mu^2}$$

indicates that the two types of governors should exhibit identical frictional characteristics. Fig. 12 represents the theoretical plot of dial speed in pulses per second versus coefficient of friction. The single curve satisfies both the fly-bar and drive-bar governors as specified in Tables I and III. This curve shows the change in speed of a dial initially adjusted to 10.0 pulses per second, operating at normal torque, as the coefficient of friction varies. It indicates that if there is a decrease in the value of μ from that which existed at the time of initial adjustment, there will be a corresponding increase in the dial speed. The curve is drawn with $\mu = 0.25$ as representing the normal stud-to-case condition and a normal governor input torque of 7,500 dyne-cm.

The experimental data were compiled for the following operating con-

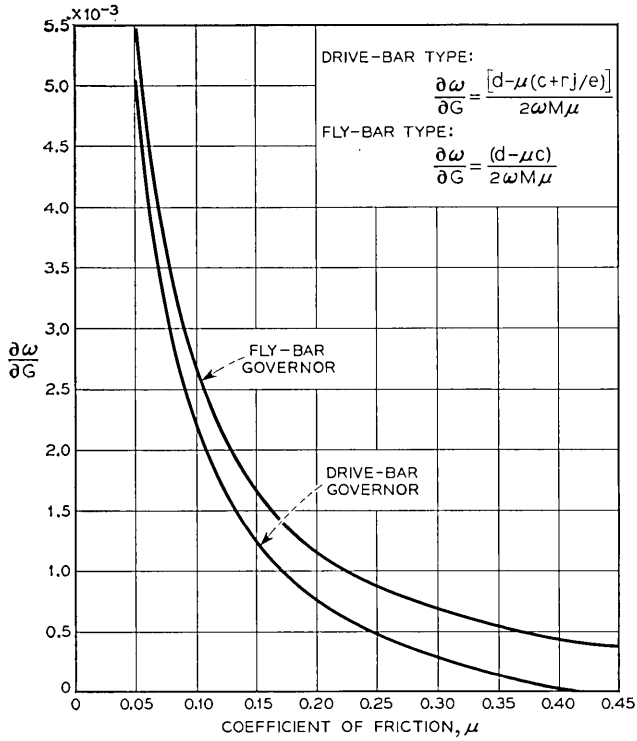

Fig. 11 — Derivative of governor speed with respect to governor input torque versus coefficient of friction.

ditions: as received, governor case cleaned with acetone, damp atmosphere, and SAE 10 petroleum base oil in the case. Each of the governors were initially adjusted to 10 pulses per second with the governor cases in the "as received" condition. The governors were then removed and the cases and governor studs were cleaned with acetone. The governors were then reassembled and the new speed recorded. During this procedure extreme care was exercised so as not to disturb the governor spring adjustment. Speed was next recorded for the damp atmosphere condition, and finally for the condition with one drop of SAE 10 oil in the governor case. For these last two conditions the governors were not removed from the dials.

The average speeds recorded for the four conditions are plotted on the theoretical speed curve of Fig. 12. The points were arbitrarily placed on the theoretical curve. No attempt was made to determine the exact coefficient of friction values corresponding to the four conditions. In this

respect, the speed attained with oil in the case shows a value of $\mu = 0.06$ which is very close to the $\mu = 0.05$ taken as the lower limit.

To produce a decrease in governor speed one must increase the value of μ. This is difficult to do on a controlled basis, since it is brought about by the progressive action of wear particles and foreign material scoring the surface of the brass case during the life of the governor. The theoretical analysis indicates that when the coefficient of friction increases to $\mu = 0.35$ the speed of the governor will decrease from 10.00 to 9.65 pulses per second. This assumes that there would be no speed change due to wear in any part of the governor. In practice of course the governor studs wear as the case is scored and, therefore, are made progressively shorter. For this condition the increased outward motion of the weight required for stud-to-case contact produces an increase in the spring force, F_s, acting on the weights. The speed change which results from increasing the spring force is in the direction to compensate for the loss in speed due to increase in coefficient of friction. Therefore, considering only stud friction and wear as effective in causing change in speed, generally the governor and dial speed should increase from its initially adjusted value during life.

It can be concluded from the experimental and theoretical evidence that there is the possibility of a speed change due to varying coefficient

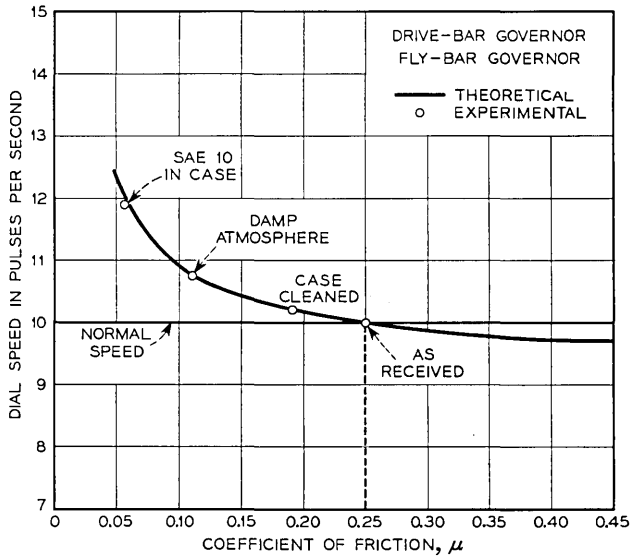

Fig. 12 — Dial speed versus coefficient of friction.

of friction to values between 9.65 and 12.0 pulses per second during life for dials initially adjusted to 10.00 pulses per second. An increase in speed may result from a decrease in stud friction due to the existence of lubricant or high humidity in the stud operating region or, the speed may decrease due to high friction.

The above changes represent the extremes in dial speed determined solely by reaction of the governor to change in stud friction. In practice it is anticipated that dials adjusted to 10 pulses per second initially can vary from 9 to 11 pulses per second during normal usage. A reduction in speed will occur as more torque is required to compensate for the increase in bearing friction caused by the accumulation of dirt and wear products during ordinary life. This additional bearing drag will cause a decrease in the torque available for governoring and therefore the dial will be regulated at reduced speed. For extreme cases of wear and contamination, it is of course possible that the dial will stop altogether during run-down. Such cases are not controllable by the governor. They result from the expected attrition during extended life or unfavorable environment.

To guard against excessive increase in dial speed from the value at time of initial adjustment, precaution is taken during manufacture. As stated previously, lubricant traveling into the governor case after initial adjustment will cause a marked increase in dial speed. To avoid this sort of contamination, the governor case is washed after machining and swabbed with clean chamois prior to insertion of the governor onto the shaft. Care is also taken to see that no lubricant is placed in the governor case during lubrication of the shaft bearings. These practices assure that initially the friction surfaces are relatively free from contamination. The increase in dial speed up to the 11 pulses per second possible during dial life will result primarily as a result of stud wear, increase in efficiency of the mechanism, and operation during periods of high humidity.

Dial Speed Versus Governor Input Torque

To substantiate the theoretical conclusion that the drive-bar governor should exhibit better regulation due to changes in input torque, experimental data were compiled on the dials equipped with the two type governors. The results of this test are plotted on Fig. 13, along with theoretical forcing curves for both governors. A coefficient of friction value of 0.25 was assumed in arriving at the theoretical curves.

The dials were initially adjusted to 10.00 pulses per second by bending the governor spring to have proper tension. Loads of 1, 3 and 5 lb were

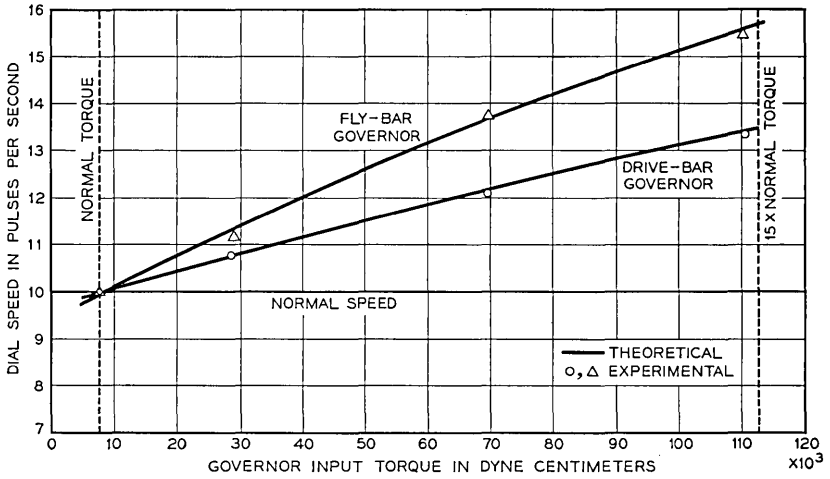

Fig. 13 — Dial speed versus governor input torque.

suspended from the fingerwheel at 5/8" radius and released. Average experimental speeds were recorded for the three forcing conditions and are noted in Fig. 13 for both the fly-bar and the new drive-bar governors. Good agreement, between the theoretical and experimental values, is evident. For a forcing condition of fifteen times normal motor spring torque, an average speed of 15.6 pulses per second is shown for the fly-bar governor while an average speed of 13.4 pulses per second is noted for the drive-bar type. Theoretically the speeds should be 15.3, and 13.2, respectively. This type agreement is also present for the 1 and 3 lb forcing conditions, and therefore, it may be concluded that for any input torque resulting from forcing the fingerwheel a dial equipped with a drive-bar governor will exhibit less speed increase than one having the fly-bar governor.

The theoretical analysis indicates that for the torque available during normal rundown, drive-bar governors as specified in Table I will decrease in speed from 10.00 to 9.80 pulses per second and fly-bar governors as specified in Table III will decrease in speed to 9.70 pulses per second. These theoretical speed changes were checked experimentally by recording on a rapid record oscillograph a trace of the make and break times of the pulsing contacts during rundown of the dial from digit zero. This information was used to determine the average dial speed in pulses per second for each sequence of make and break times. Actual loss in speed from the first to the ninth pulse for dials equipped with drive-bar gover-

nors averaged 0.24 pulses per second while dials with fly-bar governors decreased 0.33 pulses per second. The slight additional loss in speed noted experimentally was probably due to friction in the gear mechanism which is not considered by the theoretical analysis. However, since the differences recorded are quite small, one can conclude that the theoretical and experimental results are in good agreement even when concerned with small changes in torque experienced during normal rundown of a dial.

CHATTER IN GOVERNORS

It is not uncommon for governors with fine regulating ability to produce an objectionable chattering noise when operated near or at the vertical position. This chattering, while in most cases not particularly adverse from a regulation or wear point of view creates in the mind of the listener grave doubts as to the correctness of the design. In severe cases a sharp noise is heard during every half revolution of the governor shaft as each weight alternately leaves the case and strikes against the end of the other governor weight. During every revolution of the governor shaft, each weight is alternately supported as shown in the schematic, Fig. 4. At this instant, the gravity moment about B is a maximum, and along with the spring moment, opposes the centrifugal moment. If the gravity component, or effective mass of the weight, is sufficiently large, a new system is produced which has a critical velocity in excess of the regulated speed. Since the governor speed is continually regulated by the bottom weight at a speed lower than the new critical velocity, the top weight falls from contact with the case.

The magnitude of the gravity component is a function of the angle at which the governor operates and is a maximum when the dial is in the vertical position. As the operating plane of the dial decreases to the horizontal, the gravity effect decreases to zero. Chatter will not occur when the operating angle produces a gravity component smaller than the difference between the centrifugal force and the spring force.

Since the chattering effect is the result of a balance of forces on the governor, it is apparent that a relationship can be derived which will express the effect in terms of governor constants. This derivation is given in Appendix III and shows the chatter equation for a conventional fly-bar governor to be

$$m \leqq \frac{G(d - \mu c)}{2r\mu\ell \sin \beta} \tag{19}$$

This expression must be satisfied if the governor is to operate free of chatter.

By substituting the constants for the fly-bar governor, in Table III, we have for this governor operating in the vertical plane

$$3.9(980) \leqq \frac{7{,}500(0.390 - (0.25)(0.361))}{2(118)(0.25)(1.092)}$$

or

$$3{,}820 \text{ dyne-cm} \leqq 2{,}790 \text{ dyne-cm}.$$

Since the equation is not satisfied instability should be present and governors of this fly-bar design do chatter loudly when operating in the vertical plane.

The chatter equation for the fly-bar governor indicates that by adjusting the design constants, one can eliminate the instability effect. This is true. A set of values could be used which would result in a fly-bar governor which operates free of chatter. Unfortunately such a governor would also have reduced ability to govern. The relationship between chatter and governing is explained as follows.

The equations which define changes in governor speed with respect to changes in friction and torque for the fly-bar governor

$$\frac{\partial \omega}{\partial \mu} = -\frac{Gd}{2\omega M \mu^2}$$

and

$$\frac{\partial \omega}{\partial G} = \frac{1}{2}\frac{(d - \mu c)}{M\mu\omega}$$

and the chatter equation, (19), show that operation without chatter and good speed regulation are totally incompatible. Those terms in the equations which should be small for good speed regulation; i.e., torque G and stud location d and c, should be large to avoid chattering of the governor. Those terms which should be large for good speed regulation; i.e., case radius r, friction μ, and the distance from the pivot to the center of gravity l, must be small for no chatter. As the theoretical analysis indicates there is no term in the fly-bar governor chatter equation which can be operated on to eliminate chatter without impairing regulation of speed.

A similar analysis, given in Appendix IV, shows the chatter equation for the new drive-bar governor to be

$$m \leqq \frac{G(d - \mu c - \mu r j/e)}{2r\mu l \sin \beta} \tag{20}$$

This expression must be satisfied if there is to be no chattering during operation. A comparison of this equation with that determined for the fly-bar governor, equation (19), shows an additional term $(\mu rj/e)$. Substitution of the design constants for the drive-bar governor given in Table I leads to the following.

$$3.9(980) \leqq \frac{7{,}500(0.390 - (0.25)(0.361) - (0.25)(1.18)(0.236)/(0.498))}{2(1.18)(0.25)(1.092)}$$

or

$$3{,}820 \text{ dyne-cm} \leqq 1{,}862 \text{ dyne-cm}.$$

This indicates that instability should be present in this governor, and that the additional term $(\mu rj/e)$ causes a greater difference between the mass term and the torque term, than for the conventional fly-bar governor. This is to be expected, since for the drive-bar governor also, adequate speed regulation, and a design which has no chatter, are totally incompatible. The sensitivity of speed to torque change and changes in coefficient of friction.

$$\frac{\partial \omega}{\partial G} = \frac{1}{2} \frac{(d - \mu c - \mu rj/e)}{M\omega\mu}$$

and

$$\frac{\partial \omega}{\partial \mu} = -\frac{Gd}{2M\omega\mu}$$

were made as small as possible for best regulation and this results in a small value to oppose chatter.

This chatter analysis indicates that a new approach in design is needed in order to provide a governor which will operate without excessive noise and still regulate speed as required for use in the 7-type dial. This is found in a design which prohibits rapid movement of the governor weights during the unstable period.

Referring to the assembly drawing of the drive-bar governor, Fig. 3, which shows the governor in the rest position, one can see that each governor weigh rests on the end of one of the arms of the drive-bar. By supporting the weights in this manner the following two beneficial effects are achieved. One, during operation in new assemblies, the weights move outward to touch the case only a nominal distance of 0.007″. This small allowable motion in the drive-bar governor results in a low velocity of the weight at closure and hence, less impact noise. Two, because the drive-bar presses to rotate the governor weights, impact occurs as the

unstable weight skids against its arm of the drive-bar. This introduces friction damping to still further reduce the noise on closure of the weight. Some additional damping is provided by making the drive-bar and weight of powdered metal rather than wrought material. These features, which result from the particular physical arrangements of the component parts make possible the relatively quiet operation of the drive-bar governor. Experience with dials equipped with drive-bar governors indicates that they are effective since the noise due to chattering has been satisfactorily reduced so as to not be objectionable.

SUMMARY

This study has carried forward the work of C. R. Moore by presenting the derivation of theoretical equations which define speed for the drive-bar type governor. Design considerations necessary for optimum speed regulation indicated by the theory were applied in establishing the shape and working relationship of various components in the drive-bar governor. Governors constructed to these dimensions have operated as forecast by the theory. The excellent agreement between theory and practice indicates that it is both desirable and practicable to apply the Moore theory in the design of governors.

The initial requirements imposed on the design of a governor for the 7-type dial were two-fold. The new governor had to provide speed regulation at least equal to the conventional fly-bar type and no objectionable noise could be created by the governor during operation. To better understand the reasons for noise in governors, suitable theory was developed for investigating this phenomenon. Application of this theory to any type governor results in an equation which defines chatter in terms of the constants of the governor. This equation and the equations determined by the Moore theory for speed regulation indicate the existence of an interrelationship between speed regulation and noise in governors. The theory indicates that noise free operation and good regulation are totally at variance. The fly-bar governor supports this conclusion since this governor having fine control of speed also produces a chattering noise during operation. To satisfy both requirements, good regulation and quiet operation, it is first necessary to design a governor which will regulate properly and secondly, if the constants selected indicate that chattering will occur, prohibit excessive noise by providing means for restraining the system during the unstable period.

This method of attack was taken in the design of the new drive-bar governor. By applying the Moore theory, a governor was developed for

use in telephone dials which is an improvement over the conventional fly-bar type governor. Fine regulation is provided by the drive-bar governor for a given change in the coefficient of friction between the studs and case. This is achieved by locating the studs as close to the weight pivot as manufacturing techniques will permit. Improved speed regulation is provided for varying input torque in the new governor as compared with the fly-bar governor. The experimental data shows the new design able to control speed approximately twice as well. It is an effective non-forcing governor, prohibiting excessive increase in dial speed as a result of forcing the fingerwheel during rundown. This nonforcing feature is achieved by applying the driving torque to the weights at a point to develop a moment about the weight pivot. This drive-bar moment assists the centrifugal force in maintaining pressure of the friction stud against the case for friction governoring.

Having established a design which provided the degree of dial speed regulation considered necessary, it was then possible to investigate the second requirement of noise free operation. Application of the chatter theory to the drive-bar governor indicated the design to be unstable. This situation was controlled by using the ends of the drive-bar to limit the fall of the governor weights. This configuration of parts allows only small movement of the weights during the unstable period and provides damping as the weights close on the arms. The small motion and friction damping in the assembly results in a governor which is relatively free from noise during operation. Experience with dials equipped with drive-bar governors indicate that the chattering effect has been controlled.

ACKNOWLEDGMENT

The unpublished work of C. R. Moore covering the theory of fly-ball, fly-bar and band type governors has served as a foundation for this paper and the experimental work of R. E. Prescott aided considerably in determining the final drive-bar governor design. The writer also wishes to express appreciation to Mr. Prescott and H. F. Hopkins for their helpful comments and suggestions during preparation of this paper.

Appendix I

DERIVATION OF THE DRIVE-BAR TYPE GOVERNOR SPEED EQUATION

Referring to Fig. 4, as the governor mechanism rotates in a clockwise direction each weight tends to move outward under the influence of centrifugal force and the torque force, F. The centrifugal force, F_m,

acts through the center of gravity of the weights radially from the turning center of the governor shaft. The torque force, F, is applied on the weights by the drive-bar arms. These forces are opposed by the spring force, F_s. A stud-to-case force, F_n, and a frictional component of this force, μF_n, act on the weights when the friction studs are in contact with the case. In deriving the equation of motion for speeds in excess of the critical velocity the following symbols will be used as noted on Fig. 4.

F — Force applied by torque on governor weights
F_n — Normal force of case acting on studs
F_s — Force exerted by spring when studs touch case
F_m — Centrifugal force acting at center of gravity of each weight
μ — Coefficient of friction
I_0 — Moment of inertia of the governor about center shaft
ω — Angular velocity of governor
ω_0 — Critical angular velocity at which studs just touch the case
m — Mass of each weight
r_0 — Radius to the center of gravity of each weight
r — Radius of governor case
α — Stud angle
Neg. Rotation — Rundown of governor

From the schematic, Fig. 4, taking moments about B we have

$$F_m b - F_s b + \mu F_n c - F_n d + F = 0 \tag{1}$$

collecting terms

$$b(F_m - F_s) - F_n(d - \mu c) + Fj = 0$$

$$F_n = \frac{b(F_m - F_s) + Fj}{(d - \mu c)} \tag{2}$$

The driving torque on the governor is $G = 2Fe$; the retarding torque, $2\mu F_n r$. The difference between the driving torque and the retarding torque is as follows:

$$I_0\dot{\omega} = G - 2\mu F_n r \tag{3}$$

where I_0 = Moment of inertia of the governor about the shaft center
$\dot{\omega}$ = Angular acceleration about shaft center

equating

$$F_n = \frac{G - I_0\dot{\omega}}{2\mu r} \tag{4}$$

Combining equations (2) and (4) and solving for the acceleration, $\dot{\omega}$,

$$\frac{b(F_m - F_s) + Fj}{(d - \mu c)} = \frac{G - I_0\dot{\omega}}{2\mu r}$$

$$\frac{2F_m b\mu r}{(d - \mu c)} - \frac{2(F_s b - Fj)\mu r}{(d - \mu c)} = G - I_0\dot{\omega}$$

or

$$\dot{\omega} = \frac{G}{I_0} - \frac{2F_m b\mu r}{I_0(d - \mu c)} + \frac{2(F_s b - Fj)\mu r}{I_0(d - \mu c)}$$

Substituting the following values for F_m, F_s, F

$$F_m = m\omega^2 r_0$$

$$F_s = m\omega_0^2 r_0$$

$$F = \frac{G}{2e}$$

then

$$\dot{\omega} = \frac{G}{I_0} - \frac{2m\omega^2 r_0 b\mu r}{I_0(d - \mu c)} + \frac{2m\omega_0^2 r_0 b\mu r}{I_0(d - \mu c)} - \frac{j\mu r G}{I_0(d - \mu c)e}$$

Substituting for the design constants

$$K = r_0/r$$

$$M = 2mr^2 bK$$

$$\dot{\omega} = \frac{G}{I_0} - \frac{M\mu\omega^2}{I_0(d - \mu c)} + \frac{M\mu\omega_0^2}{I_0(d - \mu c)} - \frac{\mu r j G}{I_0(d - \mu c)e}$$

or

$$\dot{\omega} + \frac{M\mu}{I_0(d - \mu c)}\omega^2 = \frac{G}{I_0} + \frac{M\mu}{I_0(d - \mu c)}\omega_0^2 - \frac{\mu r j G}{I_0(d - \mu c)e} \qquad (5)$$

This is of the form

$$\frac{d\omega}{dt} + g\omega^2 = h$$

or

$$dt = \frac{d\omega}{h - g\omega^2}.$$

Separating variables and integrating*

$$t = \frac{q}{2h} \operatorname{Ln} \frac{q + \omega}{q - \omega} + c \qquad \text{where} \qquad q^2 = \frac{h}{g} \tag{6}$$

Applying the initial conditions to solve for the constant c

$$\omega = \omega_0 \qquad \text{at} \qquad t = 0$$

$$c = -\frac{q}{2h} \operatorname{Ln} \frac{q + \omega_0}{q - \omega_0}$$

Substituting in equation (6)

$$t = \frac{q}{2h} \operatorname{Ln} \frac{q + \omega}{q - \omega} - \frac{q}{2h} \operatorname{Ln} \frac{q + \omega_0}{q - \omega_0}$$

Letting

$$A = \frac{q + \omega_0}{q - \omega_0}$$

Then

$$t = \frac{q}{2h} \operatorname{Ln} \frac{1}{A} \frac{(q + \omega)}{(q - \omega)} \tag{7}$$

and

$$\omega = q \tanh \left(\frac{ht}{q} + \operatorname{Ln} \sqrt{A} \right) \tag{8}$$

Equation (8) applies as the equation of motion for the drive-bar governor for speeds in excess of the critical speed when

$$g = \frac{M\mu}{I_0(d - \mu c)} \tag{9}$$

$$h = \frac{G(d - \mu c) + M\mu\omega_0^2 - \mu rjG/e}{I_0(d - \mu c)} \tag{10}$$

$$q = \sqrt{\frac{h}{g}} = \sqrt{\frac{G(d - \mu c) + M\mu_0^2 - \mu rjG/e}{M\mu}} \tag{11}$$

Appendix II

Governor Input Torque

In order to apply the theoretical speed equations and the chatter equations developed for governors one must determine suitable values

* Short Table of Integrals, Pierce, B. O., pp. 8, No. 50.

for the stud-to-case coefficient of friction, μ, and governor input torque, G. Experimental evidence indicates a μ of 0.25 exists during normal operating conditions for hard rubber on brass. The values for governor input torque given in Tables I and III and used in the theoretical analysis for the governors were determined as follows.

The initial torque applied to the governor for the period up to the critical velocity was calculated from oscillograph string traces. These traces were obtained by mounting on the end of the governor shaft a thin disc having 36 radial slots spaced uniformly about the circumference. Light, detected through the slots of the rotating disc on the element of a photo tube, appeared as a distorted sine wave on the photographic paper. The distance between two successive wave peaks represented 10° of rotation of the governor. By noting on the trace the time between peaks, it was possible to determine the average velocity of the governor at 10° intervals after release of the fingerwheel, or start of rotation of the governor mechanism. The complete plot of these velocities appear as the experimental speed curve on Fig. 6. Inspection of the experimental curve for the drive-bar governor shows constant acceleration immediately after release. This appears as a straight line in the velocity time curve. Using the slope of this line and the moment of inertia, $I_0 = 7.4$ gm cm^2, the initial governor torque was calculated as follows:

$$G_I = I\dot{\omega} = \frac{I_0}{t'} = \frac{100(7.4)}{55} = 13{,}480 \text{ dyne-cm.}$$

The governor torque value during normal rundown was found by first determining the governor stud-to-case force, F_n, and using this value in the equation

$$G = 2\mu F_n r$$

The fly-bar governor mechanism was used to determine the F_n force since the moment equation for this type governor contains measurable values. Referring to Fig. 7, the schematic of the fly-bar governor, the moment equation about B is as follows:

$$F_m b - F_s b - F_n d + \mu F_n c = 0$$

To solve this equation for F_n one must determine the centrifugal force, F_m, and the spring force, F_s. Using electronic flash equipment with an exposure time of $1/10{,}000$ of a second, it was possible to take distortion free photographs of the governor mechanism at the middle of the rundown. These photographs were taken with the governor in the horizontal

TABLE V — TYPE DIAL GOVERNOR INPUT TORQUE AT PULSE NO. 5

Dial No.	P.P.S.	F_m	F_s	F_n	G
1	9.89	37,820	32,820	15,320	9,050
	10.05	39,050	35,950	12,580	7,420
	9.92	38,100	34,300	11,650	6,880
2	10.02	38,920	35,230	11,340	6,700
	10.01	38,600	34,880	11,400	6,730
	9.96	38,400	35,000	10,420	6,160
3	10.03	38,950	34,100	14,850	8,770
	10.22	40,600	35,420	15,880	9,360
	10.08	39,400	35,000	13,522	7,980

Average 7,500 dyne-cm, approximately.

position, thus eliminating the gravity effect. The deflection of the governor spring measured on the photograph was used to determine the F_s force, and the governor speed, necessary to determine F_m, was taken from the oscillograph string trace. Three dials were tested, each having the same maximum motor spring torque, 490,000 dyne-cm, and clean governor cases and studs to produce an assumed coefficient of friction value, $u = 0.25$. For three 7-type dials with fly-bar governors, the experimental data given in Table V applies.

As determined, this 7,500 dyne-cm torque at the governor exists at the middle of rundown of the dial. In order to compare it with the 13,500 dyne-cm initial torque previously determined, it is first necessary to consider the effect of the motor spring. As stated previously, the torque provided by the motor spring during dial rundown decreases approximately 35 per cent from the initial value of 490,000 dyne-cm. Logically, the torque at the governor would decrease by the same percentage. Applying this factor to the torque value for the middle of rundown gives a value of 10,500 dyne-cm which can be compared with the 13,500 dyne-cm torque. A difference of 3,000 dyne-cm exists for the value of initial torque at the governor as determined by the two test methods. This remaining difference can be explained by considering the frictional losses in the dial mechanism during rundown. This analysis follows:

During rundown of the dial mechanism, a pair of pulsing springs tensioned against components on the pulsing shaft are alternately raised and lowered. This action allows contacts on the springs to open and close for pulsing in the telephone line. A portion of the input torque provided by the motor spring is required for performing this function. On Fig. 14 are plotted the torque curves for these pulsing springs as the pulsing

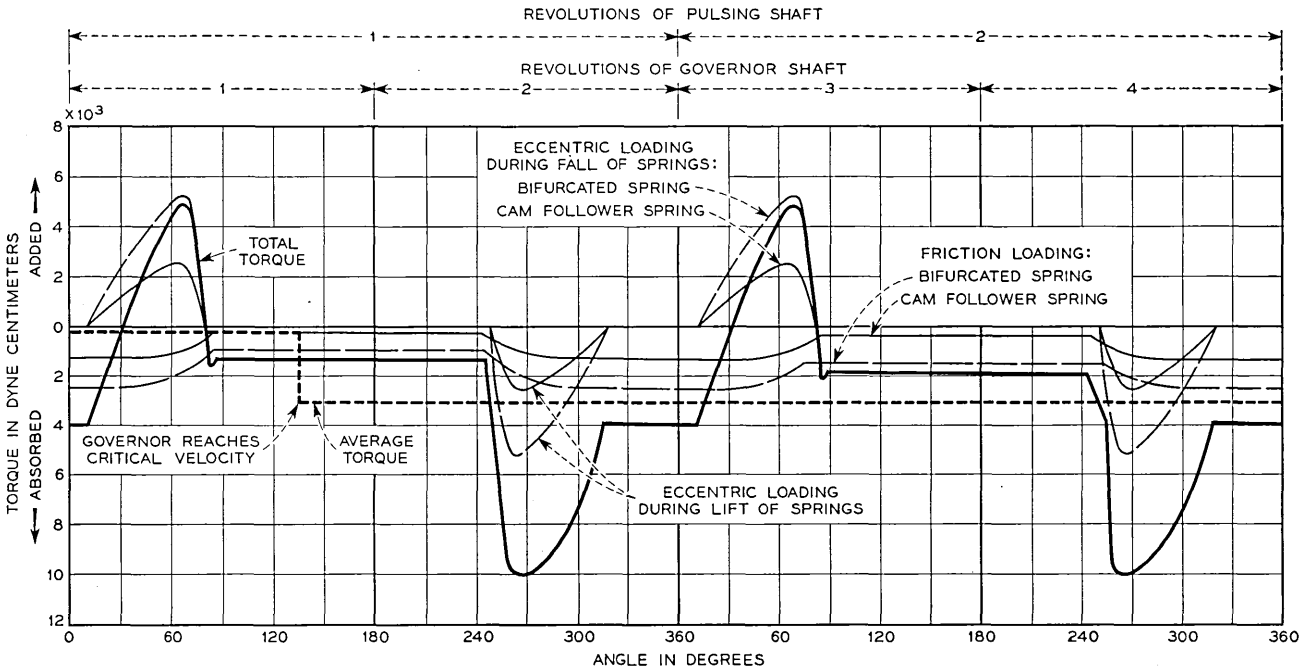

Fig. 14 — Torque analysis for the 7-type dial pulsing shaft.

shaft rotates during rundown. Fig. 15 is a schematic of the pulsing springs as they appear when the contacts are closed and when open.

For a short period during each revolution of the pulsing shaft, the pulsing springs aid the motor spring in driving the gear system. This occurs when the springs are being lowered by the cam just prior to opening of the contacts. For the remaining portion of each revolution the motor spring must provide energy to overcome frictional losses and lift the springs. These changes in energy required for moving the springs have been combined to give the total instantaneous torque curve also shown on Fig. 14.

As indicated by the pulsing spring torque analysis, the pulsing mechanism absorbs an average of only 200 dyne-cm during the period when accelerating up to the critical velocity. For rotation after the critical velocity, the average torque needed to drive the pulsing mechanism is approximately 3,000 dyne-cm. The difference between these average torque values appear at the governor shaft as a 1,500 dyne-cm torque difference. That is, 1,500 dyne-cm more torque is available for driving the governor prior to the time the critical velocity is reached as compared to that available after this time. This accounts for one half of the 3,000 dyne-cm difference in initial governor torques as calculated by the

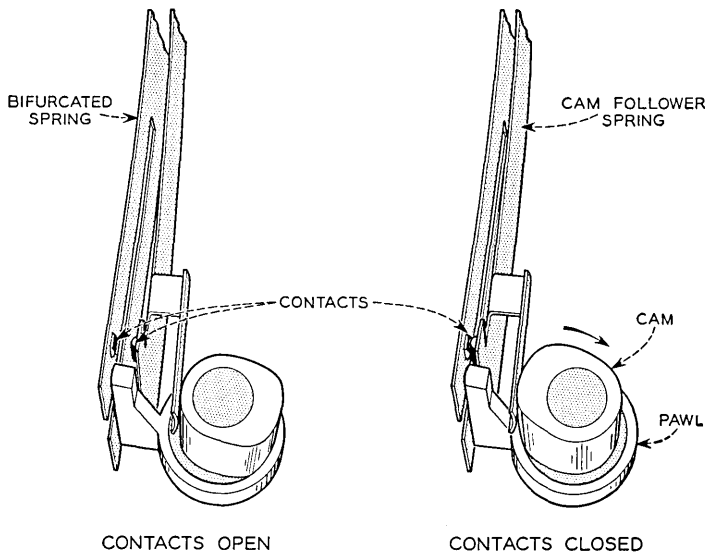

Fig. 15 — Pulsing springs of 7-type dial.

TABLE VI

Average torque at governor — Test No. 1, up to the critical velocity		13,500
Average torque at governor — Test No. 2, after reaching the critical velocity	10,500	
Torque at pulsing mechanism	1,500	
Torque required to overcome friction in the bearings	1,500	
Total	13,500 dyne-cm	13,500 dyne-cm

two test methods. The remaining 1,500 dyne-cm difference can be accounted for by considering the friction in the mechanism before and after the critical velocity.

Initially the system is accelerating as a simple fly wheel under the influence of the motor spring. At the critical velocity the governor studs engage the case and the dial rotates at virtually constant speed for the rest of the rundown period. This implies that the average speed of the moving parts in the mechanism will be twice as great for the period after the critical velocity as before. By considering the friction which exists in the dial bearings,* one can see the effect of this change in speed on the torque required to drive the mechanism. In sliding bearings with film lubrication the coefficient of friction is a function of speed. Specifically, as the speed of rotation of the journals increases, the coefficient of friction in the bearings will increase. Since the regulated dial speed is greater than the average speed while accelerating, friction in the system will also be greater. One can, therefore, justify the remaining 1,500 dyne-cm torque difference which exists before and after the critical velocity by considering it to result from the increased friction at the higher speed.

Therefore, by considering the effect of the pulsing mechanism and friction in the system, it has been possible to account for the difference in torque determined by the two test methods. Table VI shows the disposition of the torque.

APPENDIX III

DERIVATION OF CHATTER EQUATION FOR FLY-BAR TYPE GOVERNOR

Consider the governor rotating in the vertical plane at constant speed ω. Referring to the schematic Fig. 7 and taking moments about B

$$F_m b - F_s b - F_n d + \mu F_n c - ml \sin \beta = 0$$

* Design of Machine Members, Valence and Doughtie, p. 255.

where m = mass of weight
l = distance from C.G. of the weight to the pivot (B)
β = operating angle of governor weights with respect to a horizontal plane

For chattering to occur we know that the weight must leave the governor case. Therefore, at some angle β the gravity component will equal the forces tending to move the weight outward. For this condition pressure of the friction stud against the case will be equal to zero and

$$F_n d = 0$$

$$\mu F_n C = 0$$

and the moment equation becomes for this equilibrium condition

$$F_m b = F_s b + ml \sin \beta \tag{1}$$

Centrifugal Moment = Spring Moment + Gravity Moment

By applying the steady-state speed equation and the equation for centrifugal force we can transpose equation (1) to contain only design constants of the governor. From the steady-state speed equation

$$\omega = \sqrt{\frac{G(d - \mu c) + M\mu\omega_0^2}{M\mu}}$$

Substituting $F_s = m\omega_0^2 r_0$ and $M = 2mrr_0 b$

$$mrr_0 b\mu\omega^2 = \frac{G}{2}(d - \mu c) + \frac{mrr_0 b\mu F_s}{mr_0}$$

or

$$F_s = mr_0\omega^2 - \frac{G}{2}\frac{(d - \mu c)}{rb\mu}$$

also, the centrifugal force is

$$F_m = m\omega^2 r_0$$

Substituting in equation (1) the values for (F_s) and (F_m)

$$m\omega^2 r_0 b = mr_0\omega^2 b - \frac{G(d - \mu c)}{2r\mu} + ml \sin \beta$$

or

$$ml \sin \beta = \frac{G(d - \mu c)}{2r\mu}$$

where the gravity moment ($ml \sin \beta$) must be just equal to or smaller than $[G(d - \mu c)]/2r\mu$ to have no chatter occur in the governor. Expressing this in terms of the mass of the weight we have the chatter equation

for the fly-bar governor as

$$m \leqq \frac{G(d - \mu c)}{2rul \sin \beta} \tag{2}$$

Appendix IV

Derivation of Chatter Equation for Drive-Bar Type Governor

Referring to the schematic, Fig. 4, consider the governor rotating at constant speed ω in the vertical plane. Taking moments about B

$$F_m b - F_s b + F_j + \mu F_n c - F_n d - ml \sin \beta = 0$$

where m = mass of weight
l = distance from C.G. of weight to pivot B
β = operating angle of governor weights

Assume that at some angle β the gravity component will be large enough to make the F_n force equal to zero and a condition of equilibrium exists. For this condition

$$F_n d = 0$$

$$\mu F_n C = 0$$

and therefore, F_j, the torque component on the weight, must also equal zero. The moment equation becomes

$$F_m b = F_s b + ml \sin \beta \tag{1}$$

Using the equation for centrifugal force

$$F_m = m\omega^2 r_0$$

and steady-state speed for the drive-bar governor

$$\omega = \sqrt{\frac{G(d - \mu c) + M\mu\omega_0^2 - \mu rjG/e}{M\mu}}$$

where

$$F_s = m\omega_0^2 r_0 \text{ and } M = 2mrr_0 b$$

Solving for (F_s)

$$F_s = mr_0\omega^2 - \frac{G}{2r\mu b}(d - \mu c - \mu rj/e)$$

Substituting in equation (1), the values for F_m and F_s

$$m\omega^2 r_0 b = mr_0\omega^2 b - \frac{G}{2r\mu}(d - \mu c - \mu rj/e) + ml \sin \beta$$

or

$$ml \sin \beta = \frac{G}{2r\mu}(d - \mu c - \mu rj/e)$$

Expressing the equation in terms of mass of the weight we have the chatter equation for the drive-bar governor.

$$m \leqq \frac{G(d - \mu c - \mu rj/e)}{2r\mu l \sin \beta} \tag{2}$$

In-Band Single-Frequency Signaling

By A. WEAVER and N. A. NEWELL

(Manuscript received June 7, 1954)

Single-frequency signaling liberates dial systems from the restrictions of dc signaling methods. This freedom, as might be expected, is most important in the long distance telephone plant where trunks are frequently too long or have no conductors for dc signaling. The general plan of signal frequency (SF) signaling is based upon continuous signaling because of it's speed and reliability. In this respect it is like the usual dc trunk signaling schemes. SF uses steady current in the trunk signaling path for the normal idle trunk condition and no current in the signaling path for the other and alternate busy (talking) trunk condition. This choice of signal conditions is essential for SF signaling in-band systems, which as the name implies operate within the standard voice channel, to avoid conflict between signal and voice transmission. The same conditions are also used in SF out-of-band and separate line systems.

The in-band SF system can be used with any type or length of line facility that meets normal voice transmission requirements and is therefore the preferred method used by the Bell System to meet requirements for toll dialing on a national basis, with other signaling arrangements limited to the shorter trunks. The requirements, design considerations, main features, and method of operation for the in-band system are outlined in this paper.

INTRODUCTION

The signaling requirements for dial telephone operation are naturally more exacting than those for manual switching methods. This means a high order of signaling system is needed to satisfy the requirements for the toll telephone plant and for automatic toll switching systems described in recent papers in this JOURNAL.[1, 2] Indeed the advantages in speed and economy of dial telephone systems depend to a large extent upon the type of signaling provided for them. The signaling arrangements for intertoll telephone trunks which are the links between telephone switching systems, therefore, become most important.

Dial operation in the past has been based upon dc signaling which is

limited to relatively short distances and to line facilities having dc paths available to them. In planning for nationwide dialing of long-distance calls the need for an ac signaling system for dial telephone trunks became apparent. The length of intertoll trunks and the extensive growth in carrier line facilities which do not have associated dc paths made it necessary to develop ac dial signaling systems. The single-frequency signaling plan was developed for this purpose and is the first of its kind to satisfy the conditions associated with long distance intertoll dialing in the Bell System.

There are now several trunk signaling means that may be grouped as using the SF signaling plan. These are (1) the adaptation of VF carrier telegraph requiring an additional line channel independent of the voice transmission line facilities, (2) N1 and O1 carrier signaling furnished as part of these carrier terminals using 3,700 cycles outside but adjacent to the voice paths, and (3) 1600-cycle and 2400-cycle signaling systems, the in-band systems that use the voice paths.

Both in-band and out-of-band signaling have advantages and disadvantages. The in-band single-frequency signaling system uses ac in the voice frequency range to pass full supervision and dial pulsing signals over the same paths that are furnished for voice transmission in telephone trunks. This is accomplished without any loss in band width, change in line facility or addition of intermediate signaling equipment. On most calls voice and signal transmission are not required at the same time. On the few calls going to intercept operators voice transmission is impaired slightly by the effect of signal tone being on in one direction. On calls encountering busies, it is desirable to return both flashing supervision and interrupted audible tones. This can be done with out-of-band signaling but in-band signaling can return either but not both. The signaling system allows remote build-up and breakdown and provides for supervision of the temporary connections ordinarily used. Control and supervision of distant ends of trunks is required continuously whereas dial pulsing is required only at the start of calls and speech transmission is required only when connections are established.

TRUNK SIGNALS

Before going into the details of the signaling system itself it seems appropriate to review the trunk signals it is called upon to transmit. Most intertoll trunks are arranged for two-way operation, which means that a connection can originate at either end. To permit this operation, the signaling in each direction must be symmetrical and the trunk must allow the direction in which the connection is established to determine

TABLE I — CALLING TO CALLED DIRECTION

Trunk Condition	Signaling Frequency
Idle (disconnect)	On
Connect	Off
Dial pulsing*	On, then off, on pulses corresponding with dial break intervals
Ring forward	On, then off, one pulse
Disconnect (idle)	On

* Multifrequency pulsing,[3] a separate a-c signaling system, is a faster means of transmitting number information often used instead of dial pulsing. Its use eliminates only the dial pulsing signals.

TABLE II — CALLED TO CALLING DIRECTION

Trunk Condition	Signaling Frequency
Idle (on-hook)	On
Stop pulsing*	Off
Start pulsing*	On
Flashing†	Off, then on, off pulses corresponding with off hook supervision
Off-hook (answer)	Off
Ring back	On, then off, on for duration of ring
On-hook (idle)	On

* Stop- and start-pulsing control signals are required only in connection with common control switching equipment.

† Flashing supervision signals are required only for operators; the on intervals light cord circuit lamps to inform operators of the status of calls independently of the position of cord circuit talking keys.

the signaling interpretation. The latter is conveniently identified by different names for the trunk signals in the two directions. Only two signal conditions, that is, tone on or tone off, in each direction are required for all dial trunk signals. Continuous dependence upon these two conditions assures a high degree of reliability because of signal redundancy. Tables I and II show the required dial intertoll trunk signals, together with the action taken in regard to the signaling frequency.

The signaling system must be able to handle minimum and maximum length signals. The minimum times occur in dial-pulsing where the shortest signal element may be as low as 30 milliseconds. All other types of signals have longer durations.

The maximum permissible transmission time for signals between trunk terminals is determined by the allowable unguarded intervals on two-way trunks, during which double connections may occur, and also by the stop-pulsing signal recognition interval. This time is limited to about 175 milliseconds.

The distortion permitted in the transmission of signals is proportional

to the time duration of each signal. In general, variations in signal time should be within ±5 per cent. All effects of the trunk signal medium should be confined within the trunk terminals or be of such character as to have no adverse reaction in connected circuits. This is necessary for proper operation of switched connections.

BASIC PLAN

The in-band single-frequency signaling system, although fairly complex in detail, is very simple in principle. Normally, i.e., when the circuit is idle, steady tone is transmitted over the line and holds relays operated at the receiving end. Signaling is accomplished by removing and reapplying this tone, which in turn releases and reoperates the distant relays. Independent operation is obtained in each direction with one signal frequency on four-wire lines, which have separate one-way transmission pathes from terminal to terminal, and with two signal frequencies, one for each direction of transmission, on two-wire lines.

The signaling system is provided as a separate entity. It is connected in series with the transmitting and receiving branches of the line circuit at each end of a trunk and to the terminal relay circuit (trunk circuit) by two one-way signaling leads. A typical arrangement for a four-wire line terminating in a two-wire switching office at the West terminal and a four-wire switching office at the East terminal is shown in Fig. 1.

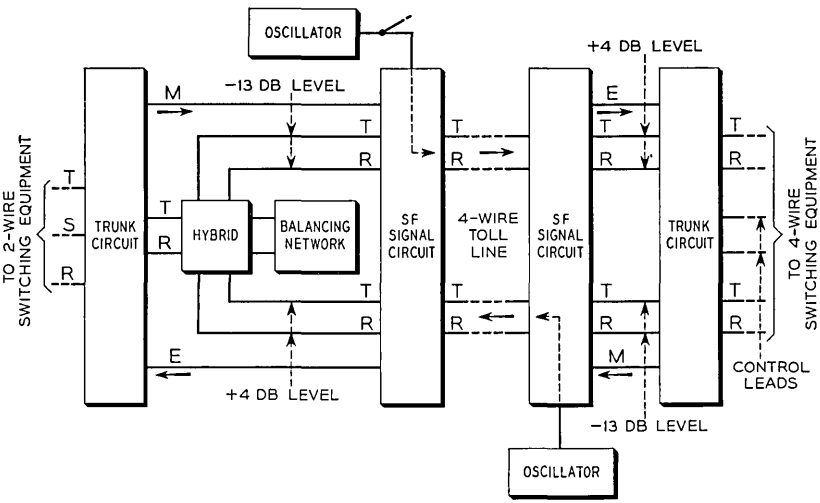

Fig. 1 — Application of single-frequency signaling to trunks with four-wire lines.

In the case of two-wire lines, the signaling equipment is applied to the four-wire transmission paths of terminal repeaters, using a different frequency for signaling in opposite directions. Band elimination networks are provided ahead of each receiving circuit to block the transmitting frequency, which would otherwise come into the receiver via echo paths and interfere with its operation.

GENERAL DESIGN CONSIDERATIONS

The successful use of the voice path for signaling, especially for continuous as contrasted to "spurt" signaling, is feasible only by a compromise among a number of conflicting factors. These factors or design considerations are (a) choice of signal frequency, (b) signaling power and receiver sensitivity, (c) imitation of signal by speech or tones, (d) interference to signal by other tones and noises, and (e) audibility of signaling tone to operators and subscribers.

(*a*) *Choice of Signal Frequency*

The choice of signal frequency is determined mainly by considerations of signal imitation by speech. As will be shown later on, signal imitation decreases rapidly as the signaling frequency is raised with the result that the highest frequency that can be reliably handled by the transmission path is used. In the case of some four-wire type lines, such as EB carrier, the highest frequency that should be used is 1,600 cycles. However, the use of 1,600 cycles results in an expensive signaling system and it is desirable to have another system using a higher frequency (2,600 cycles) for application to lines that can handle this frequency. These systems are basically the same in principle and both are described in the present article.

For application to two-wire lines the second frequency used is 2,000 cycles in the case of the older 1,600-cycle system and 2,400 cycles in the new 2,600-cycle system.

(*b*) *Signaling Power and Receiver Sensitivity*

To limit cross talk into adjacent voice channels and to avoid adding much signal power to the repeaters it is desirable to use the lowest practicable signal power consistent with a usable signal-to-noise ratio. A value of -20 dbm referred to zero transmission level for the steady idle tone is satisfactory for this purpose. In order to obtain an overall margin of 8 db the sensitivity of the receiver is set at -28 dbm. A higher power

is used for a short time at the beginning of each application of signaling tone to help overcome line noise and attenuation variation. This power increase is 14 db for the 1600-cycle system and 12 db for the 2,600-cycle system.

(c) Imitation of Signal by Speech or Plant Tones

An in-band signaling system requires that the receiver respond to signaling tone and at the same time be non responsive to speech formed currents. The principal design factors employed to achieve this feature are (1) the use of "guard action," (2) the employment of as narrow a bandwidth as practicable for the signal selective network, (3) the use of volume limiting, (4) the use of the longest operate time practicable, consistent with trunk signaling requirements, and (5) the use of the highest frequency that can be handled by the voice path.

"Guard action" is the principal means used in protecting the receiver against operation on speech. It consists in the use of nearly all frequencies in the voice band other than those in a narrow band centered on the signaling frequency to generate a voltage which is used to oppose that resulting from the signal frequency. The sum of these two voltages, plotted against frequency, for a typical receiver is shown in Fig. 2. A term used to specify the magnitude of the guard action is "guard-signal ratio" (G/S in Fig. 2) or just "guard ratio." The amount of guard which can be used is limited by signal to noise ratio because noise, like speech, tends to oppose operation of the receiver. A guard ratio in the range of 6 to 10 db has been found to be practicable.

Protection against signal imitation is also provided by narrowing the signal frequency band as much as practicable, since this reduces the effective operating power of voice and noise frequencies. However, the extent of this narrowing is limited since the operating bandwidth must be sufficient to allow for frequency variation in the signal supply, for carrier shift in the transmission path, for variation in the elements of the tuned circuit in the receiver and to allow for the transmission of the needed side bands of the signaling pulses.

A bandwidth of 60 to 75 cycles at the 3 db points at dialing power (about — 6 dbm at zero level) has been adopted as about the minimum that is practicable. Because of limiting and guard action the effective bandwidth is a function of input power and in the particular designs adopted approaches about 150 cycles at the just operate point.

Volume limiting is another means used to help prevent false operation on high levels of speech. The explanation of this action is illustrated in

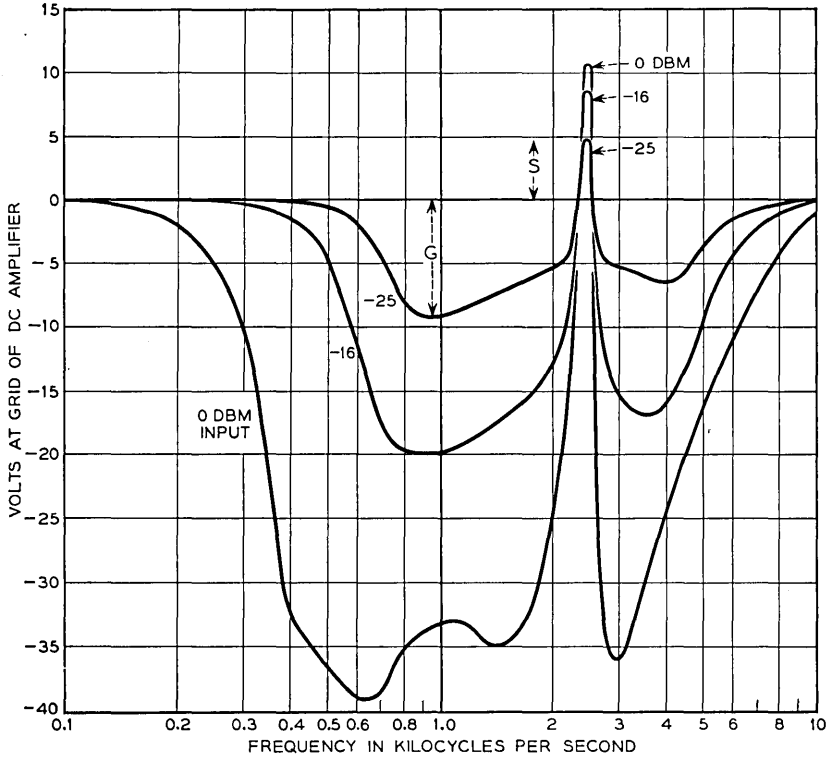

Fig. 2 — Signal-guard characteristics of 2,600-cycle receiver.

Fig. 3. The dotted line shows a characteristic for a receiver with no limiting, while the solid lines are for one with limiting. As shown, a given large input would produce an output of E_2, the difference between the signal voltage and guard voltage components, for the former case and an output of E_3, which is about half as much, for the latter case. This will be less likely to operate the receiver when applied for a short interval of time, although either will produce an operation if applied long enough, because either exceeds the just operate value E_1.

Having established the basic design parameters of sensitivity, bandwidth, guard to signal ratio, limiting characteristics and speed of response it is important to know the relationship between signal imitation and the frequency used for signaling.

To obtain information on this subject a series of tests were made using a number of guard channel receivers as nearly alike as possible except for the frequencies of signal response, which were 800, 1,350, 1,800, 2,400

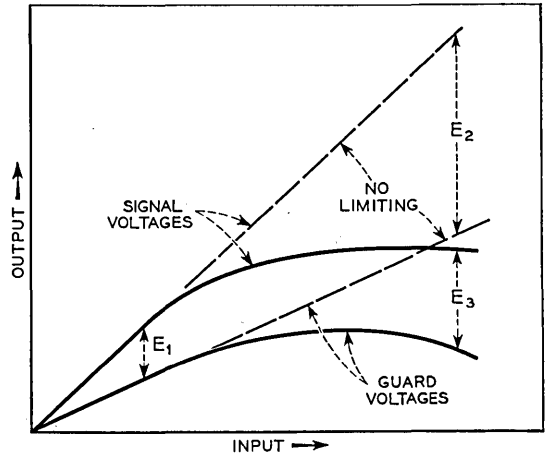

Fig. 3 — Limiter characteristics.

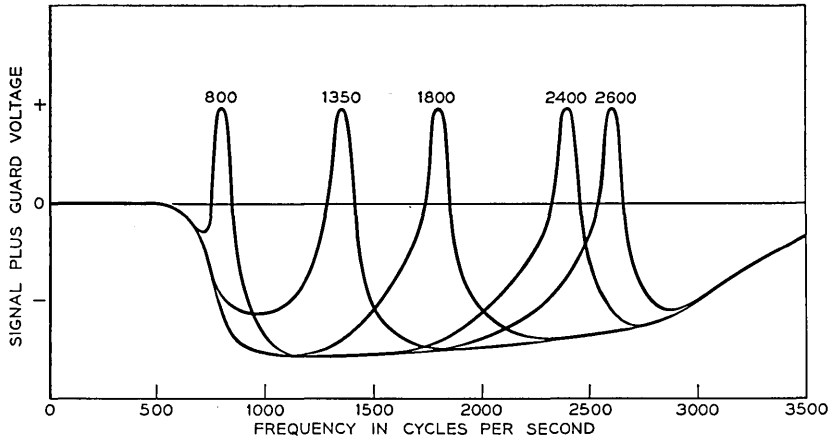

Fig. 4 — Signal-guard characteristics used in signal imitation tests.

TABLE III

Receiver sensitivity (0 level)	−28 dbm
Receiver signal bandwidth (3 db points), at just operate level approx.	150 cycles
Guard to signal ratio	6 db
Start of limiting above just operate, approx.	5 db
Minimum duration signal for operate	50 ms

and 2,600 cycles. The signal-guard characteristics of all of these receivers are shown in Fig. 4. Other parameters used are given in Table III.

The results of these tests are shown in Fig. 5, where frequency is plotted against signal imitations per 100 calls. The receivers were located at New York and were connected at different times in trunks to Boston, Toronto, Buffalo, Pittsburgh, Washington and Miami, so as to get some geographical speech distribution. There was no detectable geographic effect.

The type of speech sound causing the signal imitation is also of interest even though we have not as yet been able to put this information

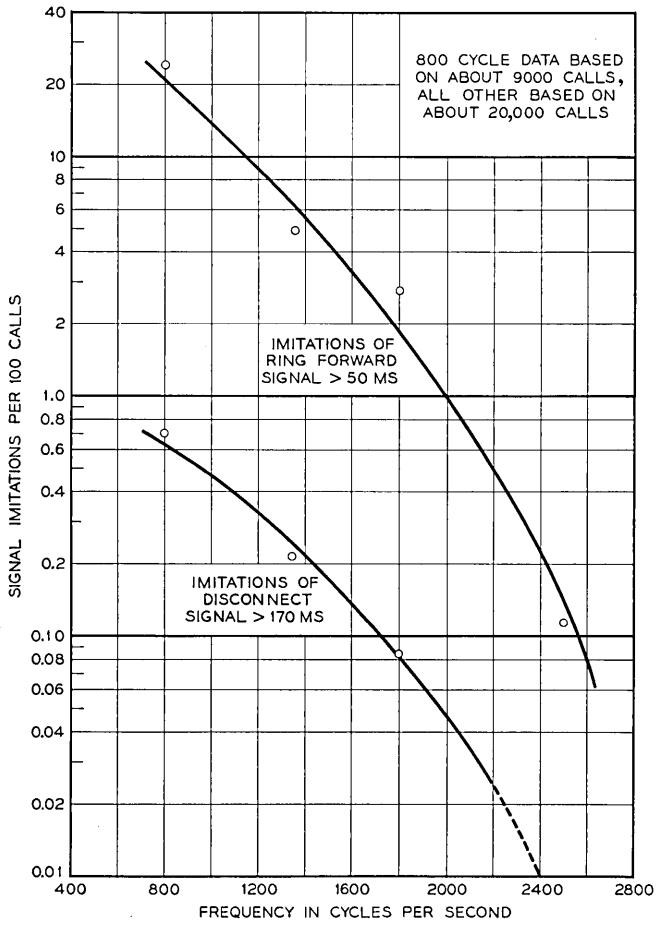

Fig. 5 — Results of signal imitation tests.

TABLE IV

Circuit "singing" or momentary circuit transient	10
Adult whistle	2
Uncertain	8
Tone spurt	1
Speech imitations	48
Total	69

to use in the design of single frequency guard type receivers. In observations on many thousands of calls it was noted that vowel type sounds were the predominant cause of signal imitations, with all except a few being formed by female speech. At the highest frequency tested (2,600 cycles) over 90 per cent were caused by the long *e* vowel sound (as in *feet*). At the intermediate frequencies 1,350 and 1,800 cycle) most vowel sounds were noted, while at 800 cycles signal imitations were caused principally by two sounds, namely *O* (as in *hole*) which accounted for about 50 per cent of the total and *aw* (as in *awl*) and similar sounds like *ah* as in *father*.

Signal imitations from vowel sounds are to be expected because of their relatively large energy and sustained nature. For instance it is well known[4] that a sustained long *e* sound can have a large component in the high frequency range with very little energy in the range from 500 to 2,000 cycles where the guard action is effective. Likewise a sustained long *O* sound can have a large peak in the 500-cycle to 1,000-cycle range with little energy in the 1,000-cycle to 3,000-cycle range where the guard action is effective for the 800-cycle receiver.

Speech formed currents are not the only source of signal imitation. In one series of observations using 2,600-cycle receivers involving circuits from New York to a number of other cities including Toronto, Boston, Baltimore, Washington and Miami a total of 69 signal imitations were observed. In each case an attempt was made to determine the sound that caused the false operation, with the results given in Table IV.

NOISE CONSIDERATIONS

Noise affects the signaling system in a variety of ways depending upon the nature of the noise and upon the particular signaling function being performed. When tone is first applied it is of course desired that the receiver operate. However at this time the "guard" circuit is functioning because it is also desired that the receiver be non responsive to speech. Noise, which acts on the guard circuit like speech, will therefore tend to prevent operation of the receiver. If the noise is steady and large enough

it would of course permanently prevent operation, while if it is of short duration and occurred at the beginning of a signal interval it would only delay operation. Even a short delay would be harmful to ring forward or dialing signals but could be tolerated in disconnect or flashing signals.

An example of how a short duration high level noise affects receiver operation is illustrated in Fig. 6. As can be seen this particular noise transient, which happened to be caused by a relay in the trunk circuit, occurred just prior to and overlapping the tone interval. It would seriously damage a ring forward signal. The solution to this particular problem is to absorb the noise at its source, or prevent it from reaching the voice path.

After the receiver is operated for a short time (0.2 sec or so) the guard action is removed. Later on when the tone is removed it is desired that the receiver release, but noise at this time will tend to prevent release. The solution to this problem is a compromise in receiver sensitivity, i.e., it must be sensitive enough to hold up on the weakest signaling tone and yet release on the maximum noise that can be tolerated from a speech point of view. Fortunately such a compromise is achieved with a sensitivity that will cause the receiver to hold with the tone about 8 db

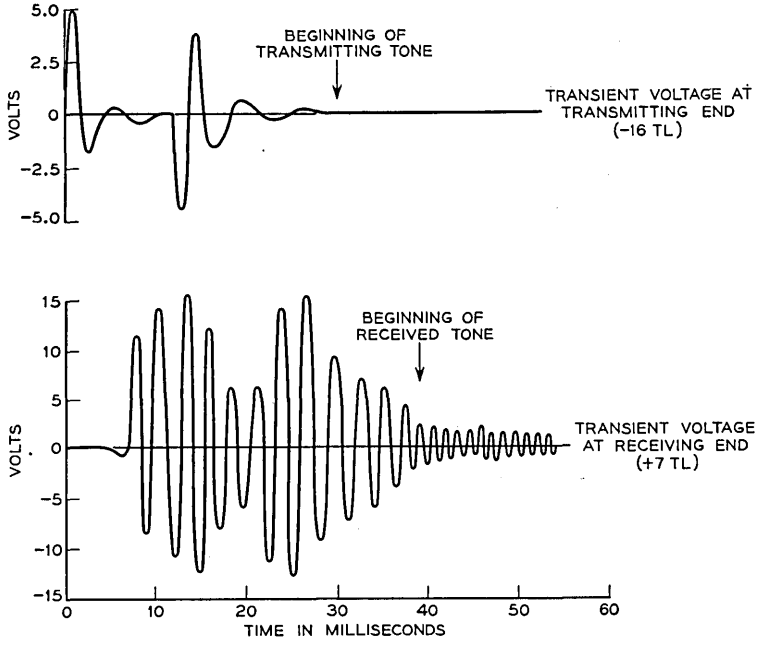

Fig. 6 — Example of transient voltages generated by relay operation.

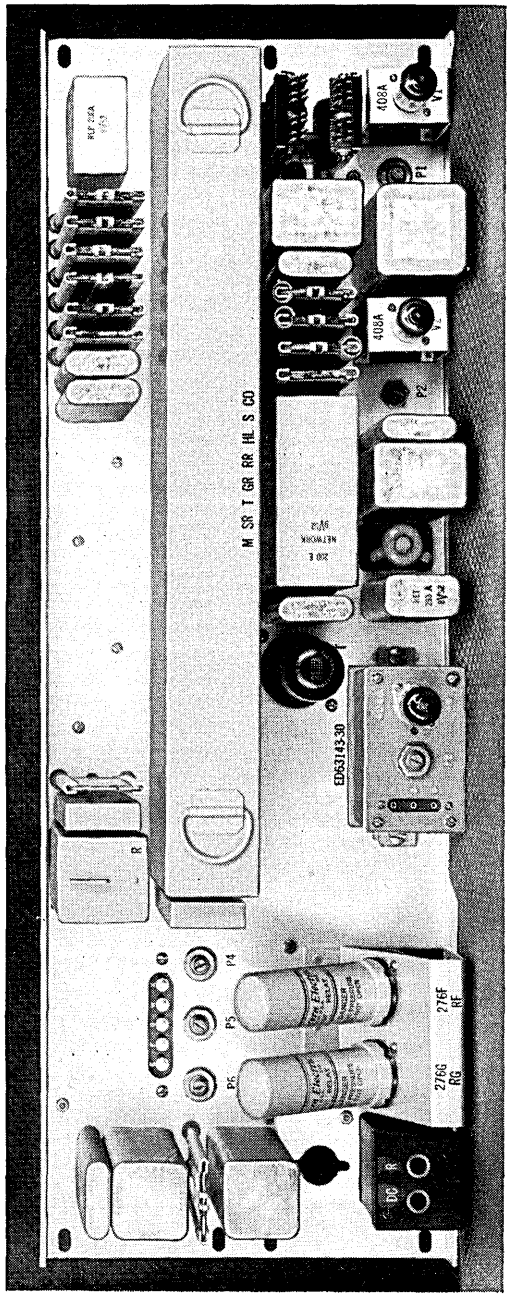

Fig. 7 — Front view of the 1,600-cycle signaling panel.

below normal and yet permit release in the presence of as much as 50 dba of thermal noise at zero transmission level.

SPECIFIC DESIGNS

The two signaling systems 1,600 and 2,600 cycles, are basically similar in principle. However, as can be seen by reference to Fig. 5, a simple guard type receiver having a frequency of 1,600 cycles would have too many signal imitations. To overcome this the guard ratio during the talking condition was increased from 6 to 10 db, the minimum signal interval to just cause a response was increased to 100 milliseconds during the talking condition and the sensitivity was decreased to −16 dbm. As a result fairly complicated timing and switching circuits are needed to assure that both transmitting and receiving circuits have the right condition at the right time.

Table V gives a summary of the principal design parameters of the two systems.

DESCRIPTION OF 1,600-CYCLE DESIGN

A front view of the 1,600-cycle main unit is shown in Fig. 7. This panel is 8 inches high by 23 inches wide and weighs about 20 lb. The essential elements of the circuit are shown in Fig. 8. It connects to the trunk relays over two leads E and M and into the line circuit via leads labeled T, R, T_1 and R_1. The transmitter, shown in the upper portion of the figure, uses dc biased germanium varistors (diodes) to control the application of signal current to the line, and for control functions, uses four relays designated M, CO, HL and RR. The functioning of the first three except for tone control are described under the heading "Description of 2,600 Cycle Design" later in this article. The RR relay (not shown) in conjunction with the M relay lengthens the sent pulse for the ring for-

TABLE V

	Dialing Condition		Talking Condition	
	1,600	2,600	1,600	2,600
Sensitivity, dbm	−28	−28	−16	−28
Bandwidth, cycles				
Low level	150	150	150	150
High level	60	75	60	75
Guard ratio, db	0	−6	10	6
Minimum signal for just operate, ms	35	35	100	50
Start of limit above just operate, db	5	5	5	5

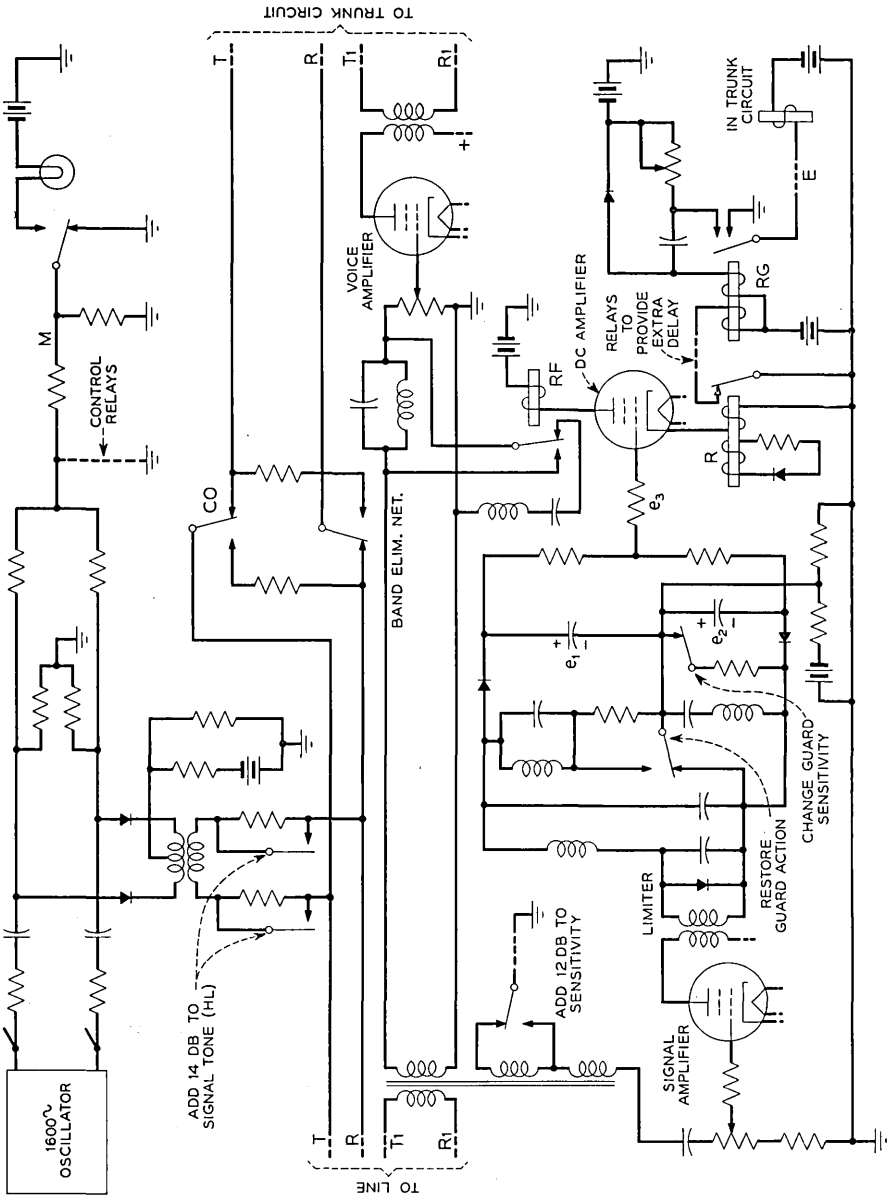

Fig. 8 — Simplified diagram of 1,600-cycle signaling circuit.

ward signal because the far end receiver at this time has a long operate time.

The signal receiver is connected in series with the receiving branch of the voice transmission path and is provided with a voice amplifier to provide a blocking function so noises originating in the switching equipment or beyond will not interfere with operation of the signaling receiver, and to compensate for the signaling bridging loss.

The receiving portion of the circuit is shown in the central and lower portions of Fig. 8. The idle condition of the trunk is shown, tone is being received, the R, RG and RF relays are operated and the band elimination filter is inserted in the receiving branch to prevent signaling tone from entering a connected circuit and interfering with signaling there.

The signal currents coming in from the line are passed through the signal amplifier, limiter and low pass filter and applied to the signal-guard network from which signal voltage is applied to the dc amplifier tube to operate the above mentioned relays and open the E lead, which extends into the trunk circuit. Typical wave forms at several points in the circuit are shown in Fig. 9. The extra operate time provided during the talking condition is obtained from slow relays (not shown) which at this time are in the path from the R to the RG relay. These relays also change the sensitivity, and guard ratio.

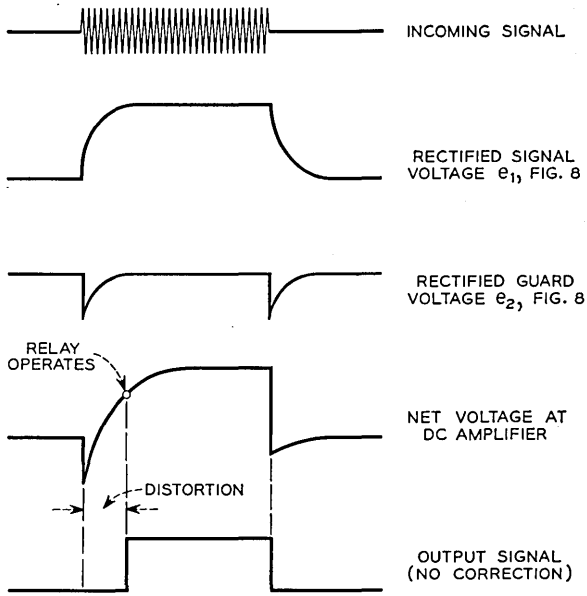

Fig. 9 — Typical wave forms in 1,600-cycle receiver.

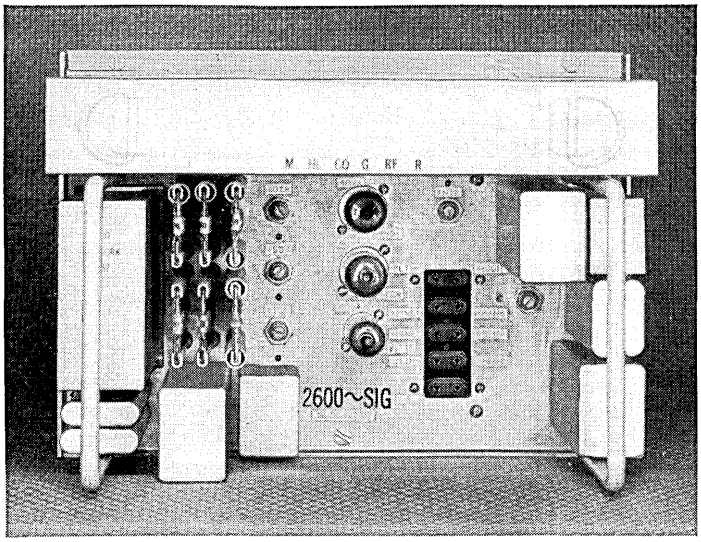

Fig. 10 — Front view of the 2,600-cycle signaling panel.

The R delay, because of guard action and the fact that its secondary winding is closed through a varistor and resistance, is relatively slow to operate and fast to release. For this reason some form of pulse correction is necessary to get good dial operation. This is obtained with the RG (regenerate) relay and its associated CR timing network, and there results an output pulse within the needed limits, even though the signal on the R relay is shortened considerably.

DESCRIPTION OF 2,600-CYCLE DESIGN

The 2,600-cycle unit, shown in Fig. 10, is just half the size of the 1,600-cycle unit, costs less than half as much, and is of the "plug in" type so it can be readily replaced for maintenance action.

A simplified diagram of the new signaling circuit is shown in Fig. 11, with the transmitting portion in the upper part of the figure and the receiver in the lower part. The transmitting portion employs three relays designated M, HL, and CO which are interconnected to perform the following functions: relay M is used to key the signaling tone; relay HL (high level) adds 12 decibels to the tone power at the beginning of each signal tone application to improve signal reliability in the presence of line noise and variations in attenuation; and relay CO (cut off) cuts the line momentarily to prevent noises originating in the switching equipment from interfering with signaling.

The operation of the receiver will be explained by describing its action (except for pulsing which will be described later) when signal frequency is received. This ac tone is amplified (or limited if it is too large) and then passed on to the signal and guard networks where a relatively high voltage results in the signaling channel and a lower voltage in the

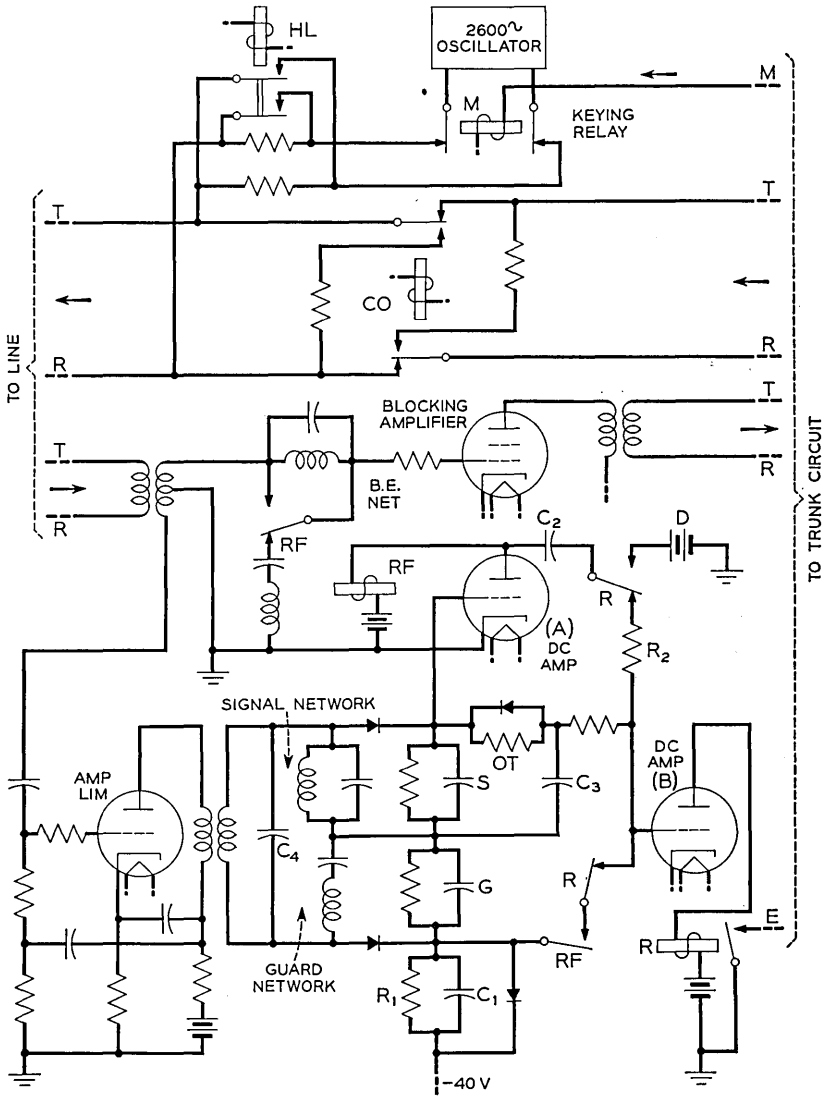

Fig. 11 — Simplified diagram of 2,600-cycle signaling circuit.

guard channel. These voltages are then applied to rectifying circuits where positive and negative dc voltages are developed and passed on to the dc amplifier tubes A and B, respectively. The RF relay operates first and cuts in the band elimination filter thereby preventing signaling tones from entering a connected toll line and interfering with the signaling there. However, a short spurt of tone will get through because of the finite time required to operate this relay, and the R relay must therefore be made slow enough so that it will not operate on this tone. This action is obtained with the resistor-capacitor network (O_T and C_3) in the grid circuit of the associated dc amplifier.

A short time (about 200 ms) after the R relay operates, a relay (not shown) releases, which short circuits the guard network and inserts enough resistance in series with the signal network to substantially remove its tuning. The purpose of removing the guard action during the idle condition is to prevent release of the receiver which would otherwise be caused by occasional bursts of line noise.

The signal network is made broad at this time for the following reason. In connections to an intercepting operator, "off hook" supervision is not returned to the originating end to avoid charging for the call. This means that tone remains on the line to continue to hold up the receiver. At the same time the intercepting operator's speech must of course be transmitted over the line so that both speech and tone enter the receiver.

Speech can be of a relatively high power as compared to the tone with the result that the action of the limiter tends to suppress the tone and could falsely release the receiver if the signal tuning were present. However, with broad tuning either speech or tone will hold up the receiver and no trouble is encountered.

The blocking amplifier seen in Fig. 11 has the same function as in the 1,600-cycle design described previously.

Among the new features in this unit, one of the most significant is the pulse-correcting circuit. This feature is a very important element in the entire long distance connection since it serves to keep the length of the dial pulses within specified time limits. The dial pulses on many calls may have to go through a number of central offices and all their associated equipment, and in each stage of the transmission path the ideal 60-millisecond dial pulse may be distorted so that it becomes too long or too short.

The pulse-correction is accomplished by generating appropriate transient voltages, whose duration is determined by capacitor-resistor networks. These voltages are then applied to the grid of dc amplifier B in Fig. 11 to perform the elongation or shortening of the pulses as required.

Dial pulses of 2,600-cycle tone with various degrees of distortion enter the circuit at the left of the diagram. It is desired that they be corrected, converted to an interrupted dc current in lead E by means of relay R, and then passed on through the toll office. This break in the dc current will effectively key the next outgoing tone transmitter, which faithfully reproduces the break intervals. The R and RF relays are shown operated as would be the case if an originally short dial tone pulse were being converted into a longer dc current break in lead E, i.e., closer to 60 milliseconds in duration.

Currents of signal frequency build up voltages across the "signal network," while currents of any other frequency, such as those from speech, build up voltages across the "guard network." These voltages are rectified separately, as shown, to obtain oppositely poled dc voltages across the signal capacitor S and the guard capacitor G. During normal speech, the voltages across the guard capacitor will dominate and, being negative, will keep the grid of dc amplifier B negative to prevent operation of the R relay at this time.

However, when dial tone pulses are applied, the main voltage will appear across the signal capacitor, with a small transient voltage appearing across the guard capacitor due to sidebands of the dialing frequency. This transient, being negative, produces an effect shown at (a)

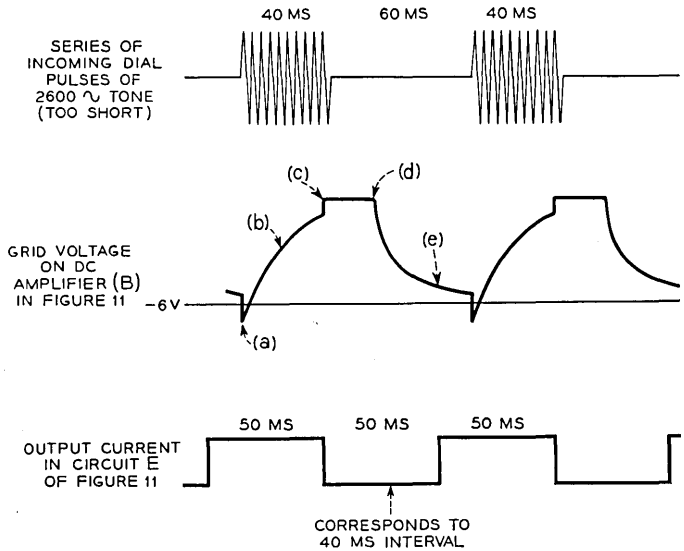

Fig. 12 — Wave forms for lengthening of short dial pulses.

in Fig. 12, the middle diagram of which represents the voltage on the grid of the dc amplifier B. The voltage on the grid of this amplifier will thereafter build up slowly as shown at (b) because of the OT (operate time) resistor and C_3 capacitor. In this illustration it is assumed that the incoming dial pulse has a duration of only 40 milliseconds, which is shorter than desired. In this particular case the RF relay operates about at time (b) and the R relay at point (c), which is just prior to the end of the pulses. It is therefore necessary, in order to obtain a corrected output signal, to delay the release of the R relay. This is accomplished by causing current to flow into the grid of the dc amplifier tube B from time (c) to time (d). This current comes from energy stored in C_2 from battery D prior to time (c) and from the positive transient at the plate of tube A generated by the end of the pulse. At time (d) the RF relay releases, and the voltage on the grid of the dc amplifier B decays along the line from (d) to (e), at which point relay R releases. The result is that the R relay has sent on a pulse that has been corrected from an original 40 to a final 50 milliseconds. It is possible but uneconomical to build a unit that would achieve perfect pulse correction to 60 milliseconds, but when a number of toll lines are in tandem, the pulses will have to pass through

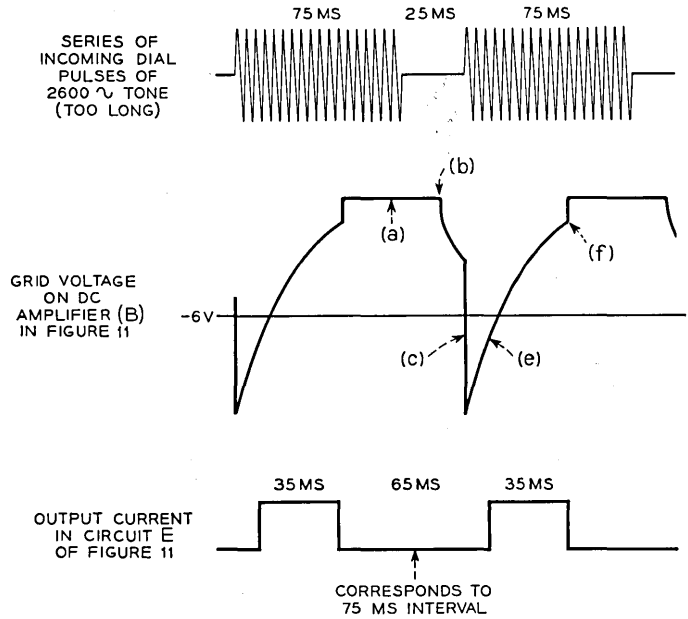

Fig. 13 — Wave forms for shortening of long dial pulses.

several SF units, and the pulses will therefore be brought close to 60 milliseconds by the successive corrections.

The action of the pulse corrector when the incoming signal is too long will be explained with the help of Fig. 13. Here the cycle of events will be assumed to start at (a), just at the end of an incoming dial pulse. From (a) to (b) current will flow into the grid of the dc amplifier B, at (b) the RF relay will release, and the voltage will start to decay, all actions so far being the same as in Fig. 12. In this instance, however, a new pulse would come along before the R relay has had a chance to release, and unless something is done about it, it would not release at all.

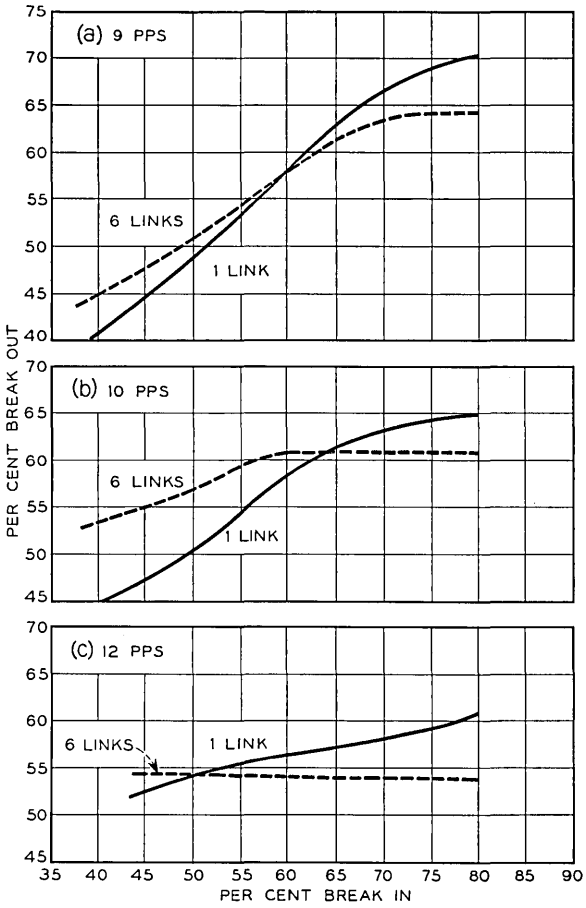

Fig. 14 — Per cent break input versus output characteristics for (a) 9 pulses per second, (b) 10 pulses per second, and (c) 12 pulses per second.

This "something" consists of using the large negative transient voltage generated at the plate of tube A resulting from the application of a positive pulse to its grid. This transient, shown at (c) drives the grid of B rapidly and heavily negative, thereby forcing the R relay to release. The remainder of the pulse-correcting action consists in using this same transient to delay the reoperation of the R relay. This is accomplished by storing some of the energy in the R_1 C_1 network. This slows the building up of the voltage as shown at (e), and the relay R reoperates at (f). The resultant repeated signal is shown below, where a 75-millisecond signal has been pulse-corrected to 65 milliseconds, further corrections being effected in the subsequent SF units.

Dialing performance of typical signaling units is shown by graphs in Figs. 14(a), (b), and (c). These curves show per cent break input plotted against per cent break output for 9, 10 and 12 pulses per second for one and 6-link operation. If the system were linear the input-output characteristic would be a 45 degree straight line. When the slope is less than 45 degrees there is pulse correction, and if the slope were zero with an output at 60 per cent, pulse correction would be perfect. It is noted that the pulse correction action improves as the speed increases and at 12 pulses per second the output is nearly independent of input.

ACKNOWLEDGEMENTS

The success of this project is the result of contributions by many people, and all cannot be named specifically. However special mention should be made of F. A. Hubbard, who designed the equipment arrangements of the 2,600-cycle system and W. W. Fritschi, C. W. Lucek, R. O. Soffel and A. K. Schenck who made important contributions to the circuit design.

REFERENCES

1. J. J. Pilliod, Fundamental Plans for Toll Telephone Plant, B.S.T.J., **31,** pp. 832–850, Sept., 1952.
2. F. F. Shipley, Automatic Toll Switching Systems, B.S.T.J., **31,** pp. 860–882, Sept., 1952.
3. C. A. Dahlbom, A. W. Horton, Jr., and D. L. Moody, Application of Multifrequency Pulsing in Switching, A.I.E.E. Transactions, **68,** pp. 392–396, 1949.
4. H. Fletcher, *Speech and Hearing,* Van Nostrand.

Centralized Automatic Message Accounting System

By G. V. KING

(Manuscript received May 6, 1954)

A centralized automatic message accounting system (CAMA) has been developed so that the billing data can be recorded at a centralized crossbar tandum office for message unit and toll calls originated by telephone customers served by a large number of local dial central offices. It is an essential part of facilities for economical nationwide customer dialing through central offices with older types of switching equipment and through other central offices which could not otherwise economically give this service. The new system records the billing data on paper tapes in the same form now used by local automatic message accounting systems. Tapes for both local and centralized automatic message accounting systems are processed in the same accounting center.

INTRODUCTION

One broad objective of the Bell System is to extend the customer's dialing range so that ultimately he will be able to dial his own calls to any telephone in the country in much the same way as he now dials local calls. Several steps toward this goal have already been taken. A revised fundamental plan[1] for automatic toll switching[2] has been adopted which involves among other things the use of a nationwide numbering plan[3] covering the United States and Canada. In accordance with this plan each customer will be given a distinctive 10-digit designation which will consist of a 3-digit regional or area code, a 3-digit central office code, and a 4-digit customer's number. In many parts of the country automatic toll switching systems[4] are now in use by operators who complete more than 40 per cent of all toll calls by dialing directly to the called telephone in distant cities.

In order that customers may use these switching systems to dial their own toll calls, some automatic means for recording the necessary billing information on such calls must be provided.

PRESENT RECORDING AND CHARGING METHODS

Several types of automatic recording equipment are now in service in the Bell System.

Multi-unit registration (zone registration) has been in use for many years in a number of panel and No. 1 crossbar central offices whose rate structure permits bulk billing. This method can be used for calls which cost 6 or less message units for the initial period. Although zone registration is economical, it does not provide a detailed record of each call but merely scores the number of message units on a register associated with the customer's line.

Remote control zone registration has been serving customers in panel central offices since 1941. It is similar to multi-unit registration, but the timing and register control equipment is located in a tandem office instead of in each originating panel central office.

Automatic ticketing,[5] which was developed some years ago for use in step-by-step central offices, does make a record of the details of each customer dialed call. A simple ticket printer is permanently associated with each outgoing trunk to produce an individual typewritten ticket for each call. Common relay equipment is used to furnish the called number, calling number, etc. to the printer. The information printed on the ticket is in detailed form and is similar to that prepared by the operator in manual operation. It can be used for billing the customer manually either on a detailed or a message unit basis.

A greatly improved form of recording, the Automatic Message Accounting[6] (AMA) system, was introduced into the Bell System in 1948. In central offices having this equipment, all of the data required for billing of customer dialed calls are automatically perforated in code on paper tapes. These tapes are taken to an accounting center where they are processed by suitable machines to produce customers' bills. The recording machines are associated with the transmission circuits only when required to make a record, one recorder serving up to 100 such circuits. Recorders, together with their associated equipment, are installed in each central office arranged for local AMA recording. The information for each call is recorded on the tape in three stages, or entries. The initial entry is recorded after the customer has finished dialing. One time entry is recorded when conversation starts and another when conversation ends. For short-haul calls that are to be billed on a message unit basis, the initial entry contains only the calling office code and telephone number, and the charging rate. This information, together with the duration of the call, is sufficient for determining the charges. On toll calls which are to be billed in detail, the called office and telephone number are also

required. This system permits individual and two party customers in No. 1 and No. 5 crossbar offices to dial calls to telephones in their home area. In addition, customers in some No. 5 crossbar central offices may now dial directly to other areas. These facilities have been installed in No. 5 crossbar offices in Englewood, N. J., and in several other locations whose customers now may dial directly to about 13 million telephones in 13 metropolitan areas.

Relatively expensive recording equipment is required in each central office in the local AMA system. For new central offices this recording equipment is economical only if the toll and message unit calling rates are relatively high. The addition of local AMA recording equipment to existing offices is, in most cases, uneconomical.

CENTRALIZED AUTOMATIC MESSAGE ACCOUNTING

The Centralized Automatic Message Accounting system (CAMA) provides an economical means of recording billing data for customer dialed calls from many central offices that cannot justify local AMA. This system is economical because one group of recording equipment, located at a crossbar tandem office,[7] can serve as many as 200 local central offices without requiring major changes in, or additions to, those offices.

The first crossbar tandem equipment arranged for CAMA was placed in service in Washington, D. C., in November, 1953. This equipment serves the customers in 85 central offices in Washington and in suburban Virginia and Maryland. They are able to dial each other directly and to dial their own calls to Baltimore and to other nearby toll points. Eventually, they will be able to dial their own calls to most points in the United States and Canada. Similar crossbar tandem CAMA equipments have been installed in Detroit, New York, San Francisco and Philadelphia.

The CAMA installation at Detroit enables the customers served by approximately 800,000 telephones in 99 Detroit panel and No. 1 crossbar local central offices to dial station-to-station multi-unit interzone and toll calls to 63 communities in Michigan and in nearby Canada. The map of Fig. 1 shows this dialing area.

THE CROSSBAR TANDEM SYSTEM

The crossbar tandem system into which CAMA has been introduced is used today in panel-crossbar and step-by-step areas. It receives calls from local dial central offices and completes them to other local central offices and to the toll network. It is also arranged to receive calls over

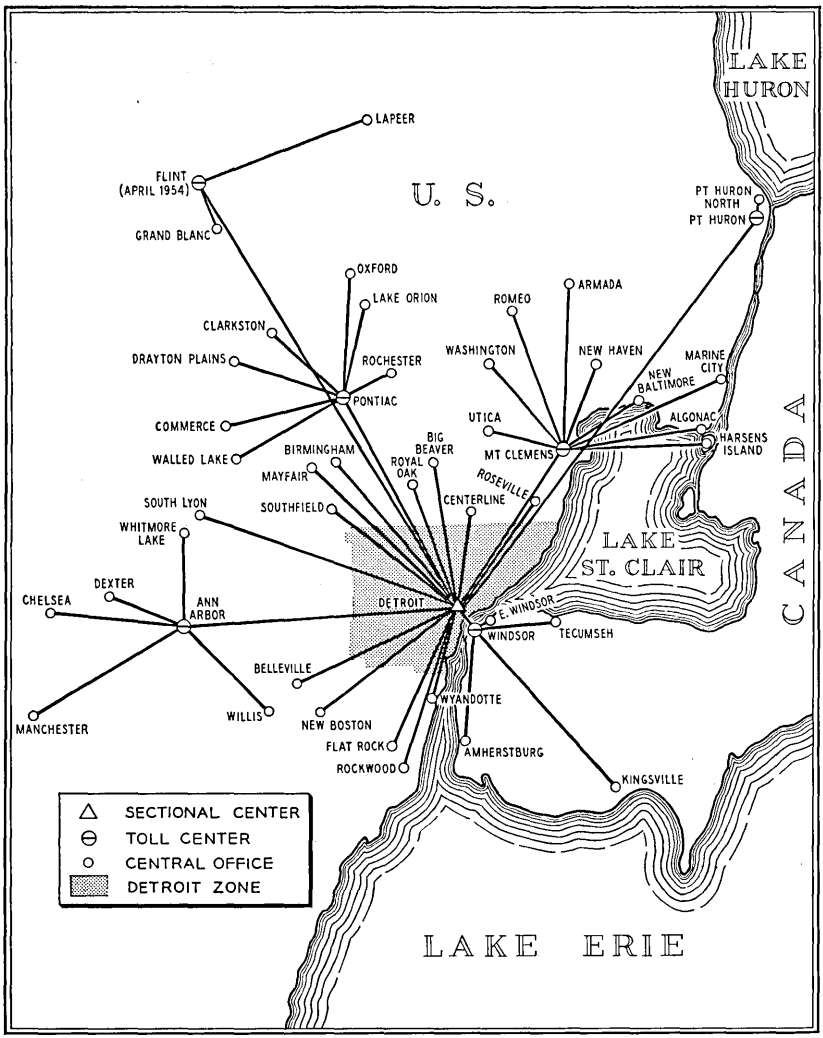

Fig. 1 — Points reached via Detroit CAMA at cutover in Dec., 1953.

intertoll trunks and complete them to other intertoll trunks or to local central offices. In many situations, it provides more efficient trunking facilities between central offices than do direct trunks and connects together offices with different signaling systems. Since the present crossbar tandem offices have in themselves no means for recording billing data, their use has been restricted to operator dialing, to customer dialing of

flat rate calls from all types of central offices, and to customer dialing of message unit and toll calls from central offices using one of the present methods of charging.

FIELD OF USE FOR THE CAMA SYSTEM

The CAMA system, as now developed, is suitable for use in panel-crossbar local areas. It is arranged to serve 7-digit calls only since facilities for 10-digit dialing are not available for panel and No. 1 crossbar central offices. Thus, in general, it can complete calls only to its own numbering area. However, provision is made for completing calls to one adjacent area, this area being selected by dialing "one-one" ahead of the listed 7-digit number.

BRIEF DESCRIPTION OF THE CAMA SYSTEM

If a customer makes a call that requires CAMA treatment, the call will be routed by the local central office to a crossbar tandem office arranged for CAMA recording. Until automatic means for identifying the calling customer's number for billing purposes is developed, an operator will be bridged on the connection at the tandem office to obtain the calling number and register it in the CAMA equipment by keying. The

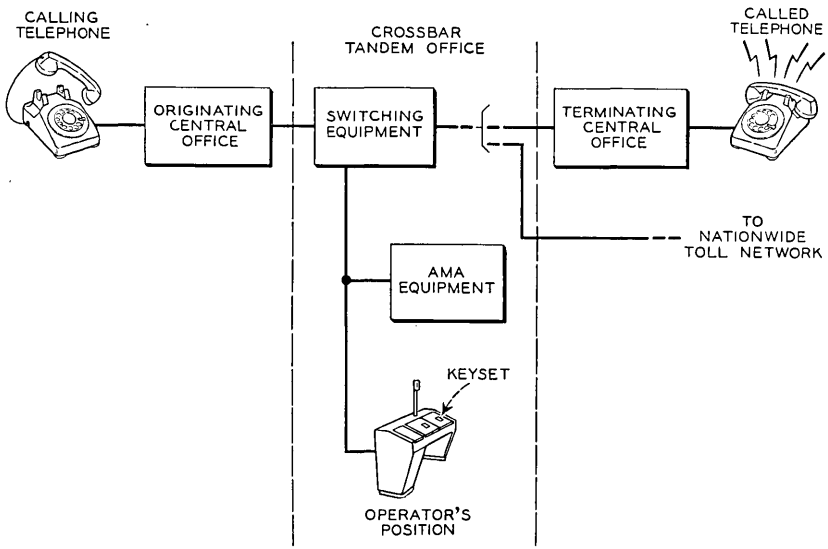

Fig. 2 – Simplified switching diagram.

information necessary for correctly charging for the call will then be recorded on paper tape by the automatic message accounting equipment located in the tandem office. This method of operation is shown in simplified form on Fig. 2.

A more detailed block diagram of the principal equipment units and their interconnections for a crossbar tandem CAMA office in a panel-crossbar area is shown in Fig. 3. Here the switching equipment consists of the conventional trunks, sender link frames, senders, markers and trunk link and office link frames for switching calls through the office. Such new units as the position link frames, positions, transverters, billing indexers, recorders, call identity indexers, and master timer constitute the major AMA equipments needed for recording the billing information for each call. Most of these have functions similar to corresponding local AMA equipments. AMA features have also been added to the trunks, sender links and senders.

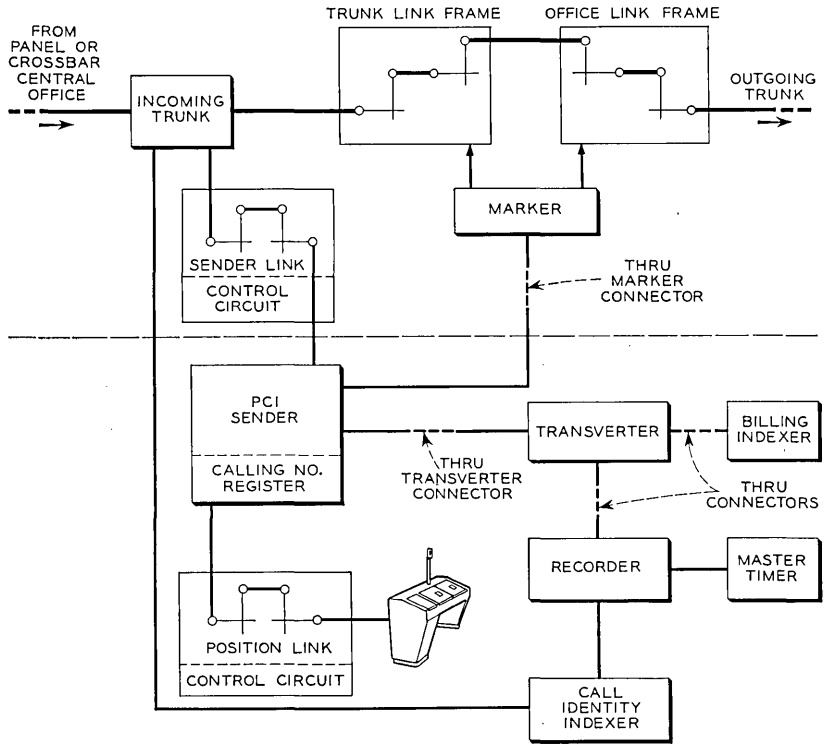

Fig. 3 — Block diagram of CAMA system in panel-crossbar areas.

FUNCTIONS OF THE CAMA SYSTEM

The functions of the system in establishing a connection and in recording the billing data may be divided into seven major groups as follows:

1. Operation of the sender link and control circuit in selecting an idle sender and connecting the selected sender to the incoming trunk.
2. Receiving and registering the called office code and number in the sender. The sender does not pulse the entire number forward until the AMA functions are completed.
3. Operation of the marker in establishing the connection through the switches and furnishing the sender with directions for completing the call.
4. Operation of the position link and control circuit in selecting an idle occupied position and connecting it to the calling customer.
5. Obtaining the number of the calling telephone verbally from the customer and keying it into the sender.
6. Connection of the sender to a transverter and billing indexer and the derivation of the billing data from the called and calling office codes and the rate class of the calling customer, and recording the charging information on the AMA tape.
7. Operation of the sender in transmitting information of the proper type to the terminating office, or if a toll call, to the next toll office in the chain.

Calls from Panel and Crossbar Customers

The first three functions in a crossbar tandem CAMA-equipped office in a panel-crossbar area are the same as in a non-CAMA office. Since published information on these features[7] is available, they will not be described in detail.

The fourth major function is handled by the position link and control circuit which is shown in block diagram form in Fig. 4. This circuit consists of primary and secondary crossbar switch links and control circuits in duplicate. Each group functions independently to serve calls to the same 40 senders and can connect to two different groups of 50 positions. In case of failure of one link group, the other link group will continue to serve calls to the 40 senders. The control circuits are arranged in such a manner that all senders and all positions receive essentially equal treatment.

When the position link connects the sender to an idle CAMA position, the operator obtains the calling number verbally from the customer and

records that number in the sender by the operation of numerical keys located at her position. The CAMA switchboard shown in Fig. 5 is a cordless board of modern sheet metal design.

The sixth group of functions — that of converting the data into the desired form and perforating it on the paper tape — is performed jointly by the transverter, the billing indexer, the recorder and the call identity indexer. The transverter is very similar to that used in local AMA. It registers the calling and called number received from the sender, registers the billing data received from the billing indexer and controls the recorder in the perforation of the initial entry.

The billing indexer is strictly a translating circuit. It receives the calling and called office codes and the customer rate class from the transverter and converts this information into a form which the transverter and recorder can use. It provides a 1-digit billing index, which denotes the charging plan to be used, and provides a type of initial entry indica-

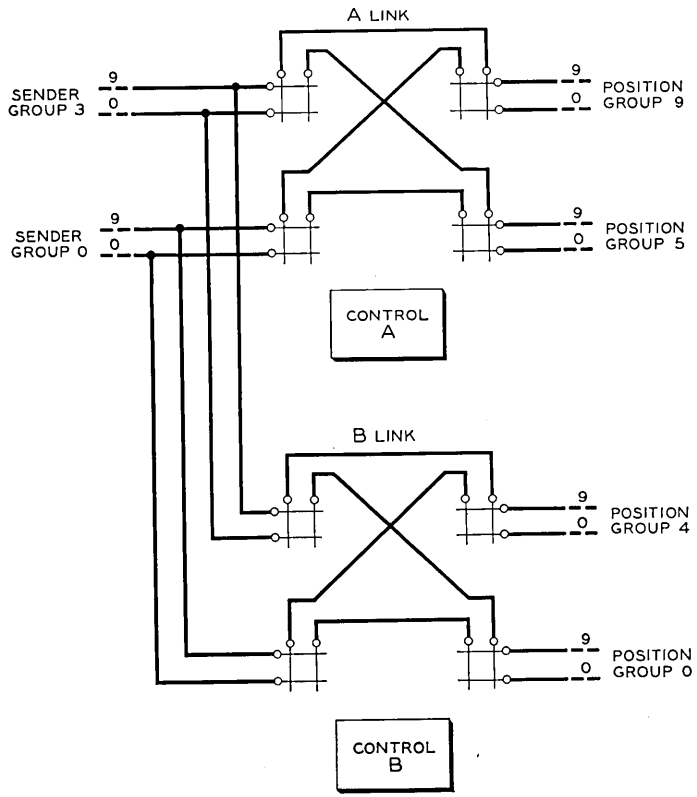

Fig. 4 — Position link and control circuit.

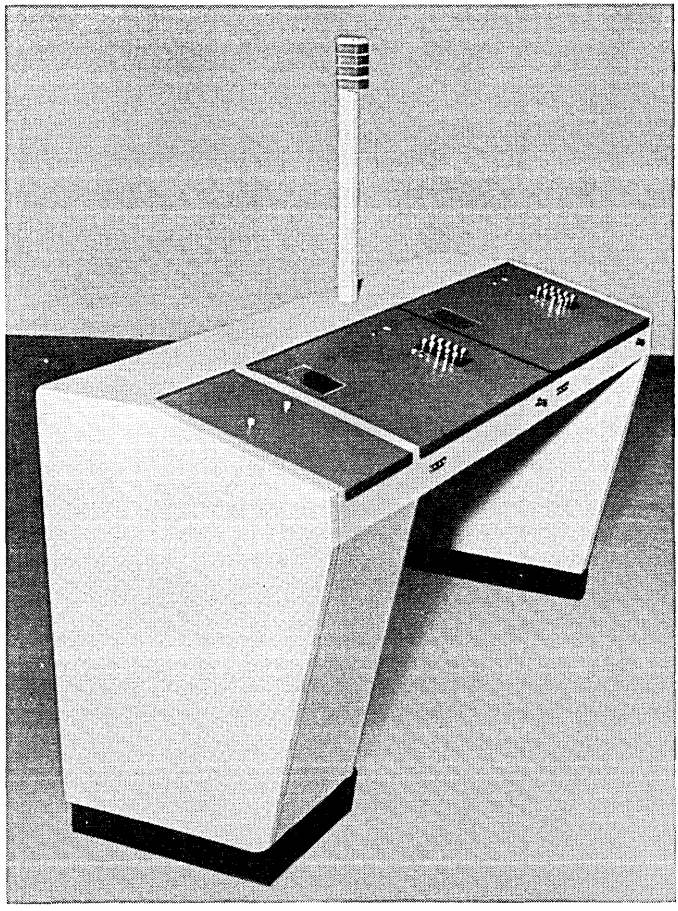

Fig. 5 — Switchboard.

tion which tells the transverter whether to perforate a two- or a four-line initial entry. The two-line initial entry used for calls billed on a message unit basis contains only the calling office code and telephone number, the billing index and the trunk identity whose function is discussed later. This information, together with the duration of the call, is sufficient for billing. Four-line entries contain, in addition, the called office code and telephone number. They are used for detail billed toll calls and for those bulk billed calls on which all details of the call are required for record purposes. Fig. 6 shows a simplified schematic of the billing indexer. The calling office code combined with the customer rate class, chooses a particular rate treatment relay. This rate treatment is common to all cus-

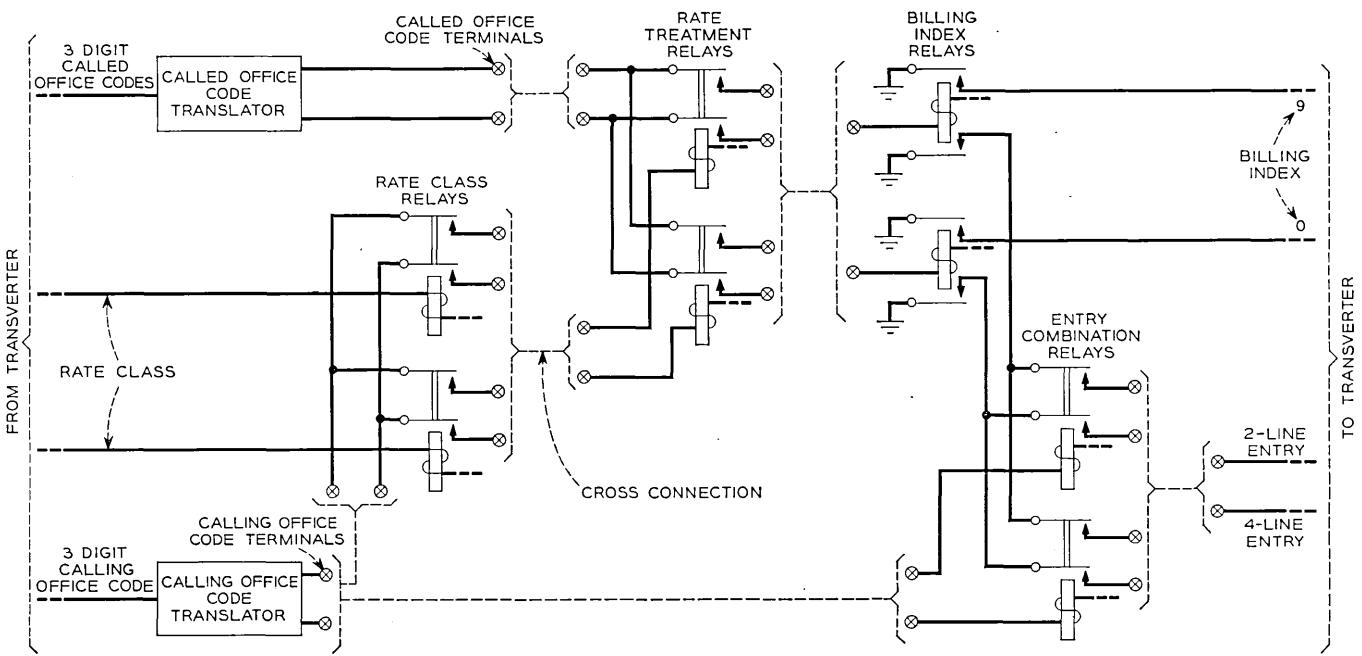

Fig. 6 — Simplified schematic of billing indexer.

tomers in the area who are charged alike for their calls through the tandem office. To actually determine the billing index or charge treatment for each call, the rate treatment is modified by the called office code by means of the rate treatment relays. Since calls with the same billing index may require a 2-line entry if originated in some offices and 4-line entry if originated in other offices, the billing index and calling office code information are translated jointly by means of the entry combination relays to produce either the 2-line or 4-line indication.

The recorder,[8] call identity indexer and master timer circuits are of the same type as used in local AMA and perform the same functions. The recorder perforates initial entries as directed by the transverter. It also perforates a timing entry at the beginning of conversation and another at the end of conversation. The call identity indexer, one of which is associated with each recorder, identifies the trunk used on a call as a particular one of the maximum 100 served by a recorder. This enables the recorder to perforate that identity on initial and timing entries. The identity is used by the accounting center to gather together the three entries involved on each call.

The master timer[9] keeps the recorders continually informed as to the correct time.

When the initial entry is completely recorded on the AMA tape, the sender completes its task of pulsing the called number forward and then releases. The transmission path is now completed through the tandem office and the only CAMA functions remaining are the perforations of the timing entries mentioned above.

ACCOUNTING CENTER PROCESS

The accounting center process for CAMA is the same as for local AMA. It automatically assembles the three bits of information pertaining to each call, computes the conversation time on all calls, sorts by the type of call, prices each call either in terms of message units for bulk-billed calls or in terms of dollars for detail billed toll calls and brings together the records of all calls made by each customer.

MAINTENANCE FEATURES

To properly maintain the AMA recording facilities, test circuits are provided for testing the major features of the CAMA equipment. An automatic incoming trunk test circuit tests the CAMA trunks. An automatic sender test circuit tests the CAMA senders in much the same way that the present sender test circuit tests non-AMA senders. Facilities are also provided for making operating tests of the position links, posi-

tions, transverters and billing indexers. As in the local AMA system, testing of the recorder features is done by the test unit on the master timer frame.

The AMA circuits are provided with many self-checking features which detect trouble while a call is being handled. When a trouble is detected, the AMA circuits connect momentarily to a recording circuit called a "trouble indicator" which lights lamps to indicate how far the call has progressed and which of the common control circuits were used on the call. This information aids the maintenance force in locating the trouble.

FUTURE DEVELOPMENTS

As stated earlier, use of Centralized Automatic Message Accounting by panel and crossbar customers will be restricted initially to calls to the home area and one foreign area. The centralized recording will be done initially at crossbar tandem offices with operators identifying the calling telephones. Ultimately, customers served by all types of dial local central offices will be able to dial their own calls — local or nationwide. Operator identification of individual and two-party lines will be replaced in many cases by automatic identification. The centralized recording equipment will be located in various types of tandem and toll offices as determined by the economics of each case.

CONCLUSION

The development of Centralized Automatic Message Accounting arrangements is another major step toward nationwide customer dialing from central offices which cannot be economically equipped with local AMA recording equipment.

REFERENCES

1. Pilliod, J. J., Fundamental Plans for Toll Telephone Plant, B.S.T.J., **31,** pp. 832–850, 1952.
2. Clark, A. B., and Osborne, H. S., Automatic Switching for Nationwide Telephone Service, B.S.T.J., **31,** pp. 823–831, 1952.
3. Nunn, W. H., Nationwide Numbering Plan, B.S.T.J., **31,** pp. 851–859, 1952.
4. Shipley, F. F., Automatic Toll Switching Systems, B.S.T.J., **31,** pp. 860–882, 1952.
5. Friend, O. A., Automatic Ticketing of Telephone Calls, A.I.E.E. Trans., **63,** pp. 81–88, 1944.
6. Meszar, J., Fundamentals of the AMA System, A.I.E.E. Trans., **67,** Part I, pp. 255–269, 1950.
7. Collis, R. E., Crossbar Tandem System, A.I.E.E. Trans., **69,** Part II, pp. 997–1004, 1950.
8. Cahill, H. D., Recording on AMA Tape in Central Offices, Bell Labs. Record, **29,** p. 565, Dec., 1951.
9. Jordan, W. C., The AMA Timer, Bell Labs. Record, **30,** p. 122, March, 1952.

The Wave Picture of Microwave Tubes

By J. R. PIERCE

(Manuscript received March 12, 1954)

Many microwave tubes make use of a long electron beam. The radio frequency excitation on such a beam can be expressed in terms of two space-charge waves, one of which has negative energy and negative power flow. The electron beam may pass through resonators, through lossy surroundings, through slow-wave circuits. In this paper the low-level operation of klystrons, resistive-wall amplifiers, easitrons, space-charge-wave amplifiers, traveling-wave tubes and double-stream amplifiers is explained in terms of waves on electron beams and on circuits. Noise is discussed in terms of such waves.

INTRODUCTION

There are many different ways in which one can make a valid analysis of the low-level or small-signal behavior of the many types of microwave tubes which use long electron beams. Which way one should choose depends partly on one's purpose in making the analysis, and partly on the particular problem to be solved.

All of these analyses lead at some point to waves or modes of propagation: waves which travel along an electron stream, along a circuit, or along the two together; waves which are unattenuated or which increase or decrease with distance. Sometimes, the analysis starts out with electron current, electron velocity and circuit dimensions as the fundamental physical quantities, just as network analysis can start out with inductance, capacitance and resistance. However, an analysis can start out instead with waves, their propagation constants and their characteristic impedances as the fundamental physical bases of the analysis. We might argue that as we are to end with waves, we may well start with waves. As it turns out, the picture of the operation of various tubes in terms of waves is simple and pleasing.

It is the purpose of this paper to present a picture of the operation of microwave tubes in terms of waves. This may be of some interest to those outside of the tube field, in that it gives an account of many recent devices. For experts in the field it can serve as an introduction to a method of analysis which is fairly recent and which may be unfamiliar.

In this analysis, certain simplifications are made. One underlying simplification is that of linearity; it is assumed that at low signal levels the behavior of the electron stream, which is inherently non-linear, can be represented by that of a truly linear system.

As this paper purports to give an accurate and useful picture of the low-level operation of microwave beam devices rather than an exhaustive discussion, some details have been omitted or passed over lightly because they seemed to be of secondary importance. Material which may be unfamiliar to workers in other fields but which is important as background is presented in appendices A to C. Various points can be pursued further in the literature. References to publications and to those responsible for various advances are not given in the body of the paper or in the appendices; they are given for each topic in Appendix D.

SPACE-CHARGE WAVES

Many microwave tubes embody a long, narrow electron beam surrounded by a conducting tube and focused or confined by a longitudinal magnetic field. At low levels of operation, the radio-frequency disturbances on such an electron stream can be expressed in terms of space-charge waves.

In these waves, two forms of energy are of primary importance: electrostatic energy associated with the bunching together of electrons, and kinetic energy, associated with differences in the velocities of the electrons. Thus, the waves may be called electromechanical; the electric energy which we associate with waves in transmission lines and waveguides is present, but the magnetic energy is replaced by kinetic energy. In circuit terms, we have an electrical capacitive element, but the inductive element is inertial, not magnetic in nature. When the electron and wave velocities are slow compared with the velocity of light, the magnetic fields produced by the electron convection current are negligible.

There may be many space-charge modes or waves in an electron stream, some with complex radial and angular variations of amplitude over the electron stream. Two waves predominate in the operation of tubes, however, and one simplification we will make is to deal with these only, and to disregard other modes of propagation on the electron stream. Appendix A discusses such a pair of waves in a simplified physical system.

We can associate with these two waves an ac electron convection current i and an ac electron velocity v. Either we can assume that the electron beam is narrow and disregard the fact that these quantities vary

across the beam, or we can deal with peak or effective values much as in the case of voltages and currents in waveguides.

These ac quantities are assumed to contain a factor

$$e^{j\omega t}\, e^{-j\beta z}$$

That is, they vary sinusoidally with time and with distance, and (assuming β to be positive) propagate in the $+z$ direction. The phase constants β of the two waves will be called β_1 and β_2. For beams of moderate charge they are very nearly

$$\beta_1 = \frac{\omega}{u_0} + \frac{\omega_q}{u_0}$$

$$\beta_2 = \frac{\omega}{u_0} - \frac{\omega_q}{u_0}$$

Here u_0 is the electron speed, ω is the operating radian frequency and ω_q is the effective plasma radian frequency.

The plasma frequency of the electron beam ω_p is given by

$$\omega_p^{\,2} = \frac{\frac{e}{m}\rho_0}{\epsilon}$$

Here e/m is the charge-to-mass ratio of the electron, ρ_0 is the charge density and ϵ is the dielectric constant of vacuum. In terms of ω_p, ω_q may be expressed

$$\omega_q = R\omega_p$$

Here R is a factor somewhat less than unity which depends on the geometry of the electron beam, on ω and ω_p, and on the velocity distribution of the electrons (see Appendices A and C).

Let us consider the simple case in which R is unity and the effective plasma frequency is equal to the plasma frequency. The phase velocities v_1 and v_2 of the two waves, which are ω divided by β, are

$$v_1 = \frac{u_0}{1 + \frac{\omega_p}{\omega}}$$

$$v_2 = \frac{u_0}{1 - \frac{\omega_p}{\omega}}$$

Thus, the first wave has a phase velocity less than that of the electrons; it is a slow wave, and the second wave is a fast wave.

Suppose we make up a radio-frequency pulse out of various frequency components of one wave. The pulse envelope generally travels with a different velocity from that of the rf sinusoids under the envelope. The velocity of the envelope is called the group velocity. The group velocity is the velocity with which a signal is transmitted. The direction of the group velocity is the direction in which causality acts (for some waves the phase velocity and the group velocity have opposite directions). The group velocity tells in which direction energy flows, and the power flow P is the stored energy per unit length, W, times the group velocity, v_g.

$$P = Wv_g$$

The group velocity is given by

$$v_g = \frac{1}{\partial\beta/\partial\omega}$$

We see that for our assumption ω_q is equal to ω_p, the group velocity for each wave or mode is u_0, the velocity of the electrons in the beam

$$v_g = u_0$$

Thus, of the two waves, the first has a phase velocity slower than that of the electrons, the second has a phase velocity faster than that of the electrons, and each has a group velocity equal to that of the electrons.

A simple discussion of power flow is given in Appendix B. In describing the excitation of the electron stream we can use the convection current i together with a quantity U which is analogous to voltage. In terms of the ac electron velocity v,

$$U = -\frac{u_0}{\frac{|e|}{m}} v$$

The real power flow P is given by

$$P = \tfrac{1}{2}(iU^* + i^*U)$$

This relation is justified in Appendix B.

For each of the two waves the voltage U bears a constant ratio to the current i; this ratio is the characteristic impedance K of the wave. We find that

$$K_1 = \frac{U_1}{i_1} = -2\frac{\omega_q}{\omega}\frac{V_0}{I_0}$$

$$K_2 = \frac{U_2}{i_2} = 2\frac{\omega_q}{\omega}\frac{V_0}{I_0}$$

Here V_0 is the accelerating voltage specifying the electron velocity u_0 and I_0 is the total beam current.

We see that the characteristic impedance K_1 of the slow wave is negative. This means that the power flow in the $+z$ direction is negative. We could also say that positive power flows in the $-z$ direction, but this may carry an unfortunate implication as to the direction in which causality acts. An example may be helpful.

Fig. 1 shows an electron beam acted on by the fields of two devices A and B. The fields in A are such as to set up the slow wave only. This travels between A and B. The fields of B are such as to just remove the slow wave entirely, so that the electron beam leaves B with no ac disturbance on it. The electron velocity u_0, phase velocity v, group velocity v_g and negative power flow $-P$ are all directed in the $+z$ direction, that is, to the right.

We must remove a power P from A to set up the slow wave. A power $-P$ flows from A to B. We must add a power P to B to remove the slow wave from the electron beam. Causality acts from A to B. To change the amplitude of the slow wave between A and B we must change the fields in A, not the fields in B.

The power flow is the group velocity times the stored energy per unit length. As the group velocity for the slow wave is positive and the power flow is negative, we see that the stored energy must be negative.

If we moved with the electrons and observed the waves, we would find that the average kinetic energy associated with the ac electron velocity was equal to the average potential energy of the electric field, and that both were positive; this is characteristic of waves in a stationary medium. The kinetic energy of the electrons relative to a fixed observer is proportional to the square of their total velocity, that is, the ac velocity plus the average velocity. The average velocity is larger than the ac velocity, so that energy terms involving the product of the average

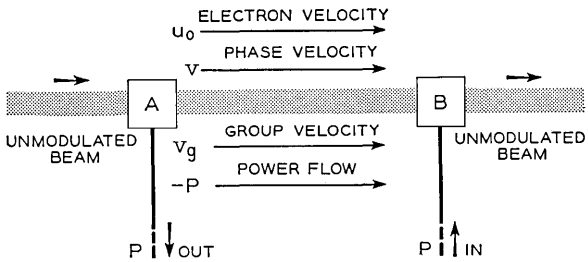

Fig. 1 — Device A sets up the slow space-charge wave only, and device B removes it. u_0, v, v_g and $-P$ are respectively the electron velocity, the phase velocity, the group velocity and the power flow between A and B.

velocity and the ac velocity are larger than terms involving the square of the ac velocity. The product terms may be negative or positive.

We can understand the negative energy of the slow wave qualitatively through a simple argument of a somewhat different sort. In the slow wave, the charge density is greatest in regions of less-than-average velocity and least in regions of more-than-average velocity, so that the electron beam has less total kinetic energy in the presence of the slow wave than it does in the absence of the slow wave. How does this come to be? Suppose that we move with the wave; we then see electrons moving in an electric field which is constant with time, and hence, as electrons move through the field their velocities vary as the square root of the potential. Relative to the wave, the electrons move slowest in the low-potential regions, and correspondingly, they are bunched together in regions of low potential. Now, for the slow wave the total electron velocity is the arithmetic sum of the wave velocity and the electron velocity relative to the wave, so if the electrons are bunched in regions of lowest velocity relative to the wave they are necessarily bunched in the regions of least total electron velocity, and the kinetic energy of the slow wave is thus negative.

In the case of the fast wave, the electrons travel backward relative to the wave. The total electron velocity is the arithmetic difference between the wave velocity and the electron velocity relative to the wave. Hence, the total electron velocity is greatest at the bunches, where the velocity relative to the wave is least, and the kinetic energy of the fast wave is positive.

THE KLYSTRON

We can explain the operation of a number of types of vacuum tubes in terms of space-charge waves. Consider the klystron, illustrated in Fig. 2. The voltage produced across the input resonator by the input signal sets up on the electron beam both the slow and the fast space-charge waves in equal magnitudes and so phased that the velocities v, or the voltages U add, while the currents cancel. Thus, just beyond the input resonator, the beam has an ac velocity; it is velocity modulated, but it has no ac convection current.

Because the two space-charge waves, one with negative power flow and one with positive power flow, are set up with equal magnitudes, the ac power flow in the beam between the input and the output resonators is zero. The input resonator neither adds power to nor subtracts power from the beam.

Because the two waves have different phase velocities, their relative phase changes as they travel along the beam. If we go along the beam a

distance L such that

$$2\frac{\omega_q}{u_0}L = \pi$$

we will find that the ac velocities of the two waves cancel and their currents add. If at this point we put an output resonator, the current will produce a voltage across the resonator which will act on the electron beam to set up new components of the slow and the fast waves.

If the resonator is on tune, so that it acts as a resistive impedance, the phase of the voltage is such with respect to the space-charge wave producing it that the new component of the fast space-charge wave

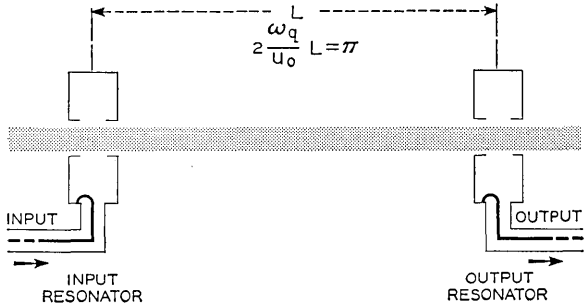

Fig. 2 — In a klystron the input resonator sets up slow and fast space-charge waves so phased that the velocities add and the currents cancel. At the output resonator the currents add and the velocities cancel. The voltage across the output resonator increases the amplitude of the slow, negative-power wave and decreases the amplitude of the fast, positive-power wave.

subtracts from the old component, while the new component of the slow space-charge wave adds to the old component. Thus, while to the left of the output resonator the two space-charge waves have equal magnitudes, so that the net power flow is zero, to the right of the output resonator the slow space-charge wave has a greater magnitude than the fast space-charge wave, so that the power flow in the beam is negative. The missing power appears as the output from the output resonator.

Of course, klystrons are frequently used in the nonlinear range of operation, and the distance L between resonators may be chosen differently from other considerations.

THE RESISTIVE-WALL AMPLIFIER

Consider a tube much like a klystron, but in which the electron beam is surrounded by a glass tube coated with lossy material, such as graphite, as shown in Fig. 3.

As in the klystron, the input resonator produces both the slow and the fast waves with equal magnitudes. As each wave travels, it induces currents in the resistive wall surrounding it and dissipates power in the wall. Thus, the power in each wave must decrease as the wave travels.

The fast wave has a positive power, and so for its power to decrease the amplitude must decrease. Thus, in the resistive-wall region the amplitude of the fast space-charge wave decreases exponentially with distance.

Because the slow space-charge wave has a negative power, its power can decrease only if the amplitude of the wave increases, so that the power flow becomes less (more negative). Thus, in the resistive-wall region the amplitude of the slow space-charge wave increases exponentially with distance; the wave has a negative attenuation; it is amplified as it travels.

If we put the output resonator far from the input resonator, the amplitude of the fast space-charge wave will be very small there, but the amplitude of the slow space-charge wave may be very large. Its current will produce a large voltage across the output resonator. As in the case of the klystron, this voltage will increase the amplitude of the slow space-charge wave, thus decreasing the power flow in the electron stream.

The resistive wall amplifier has a feature which the klystron lacks; the process of amplification involves an actual growing wave along the electron stream.

THE EASITRON; INCREASING WAVE IN A LOSSLESS SYSTEM

Consider a tube somewhat similar to the resistive wall amplifier, but in which the beam is surrounded, not by a lossy tube, but by a series of pill-box resonators, as shown in Fig. 4. Imagine that the resonators are so tuned that at the operating frequency they present a lossless negative susceptance to the electron beam.

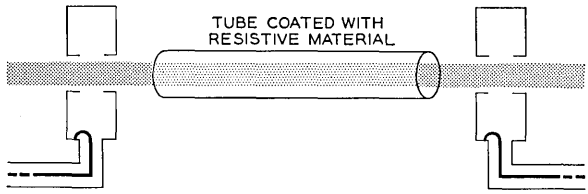

Fig. 3 — In a resistive-wall amplifier the currents excited in the lossy wall by the slow, negative-power wave decrease the power in the wave, so that the amplitude of the wave must increase.

The impedance an electron beam sees in traveling through free space or in a concentric lossless tube is capacitive. In section 1 the space-charge waves were described as involving the stored energy of the electric field, capacitive in nature, and the kinetic energy of the electrons, which has an inductive effect. We might liken the beam and its capacitive circuit to the ladder network of Fig. 5. We know that such a network supports waves.

When the charge of the beam sees a negative susceptance, the behavior is much as if the capacitances in the ladder network of Fig. 5 were negative.* In this case the waves characteristic of the circuit are not traveling waves, but are a pair of waves, one of which decays with distance and one of which increases with distance. Neither has any net stored energy.

We can express the propagation constants of the waves much as in the section on, "Space-Charge Waves," but the effective plasma frequency ω_q is now imaginary; we will call it $j\omega_q'$. The phase constants β_1 and β_2

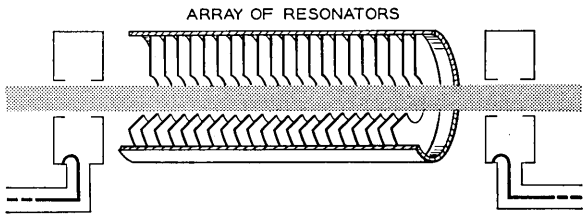

Fig. 4 — In the easitron, resonators surrounding the beam change the susceptance the electrons see from positive to negative. The system no longer supports two traveling waves, but rather, a growing and a decaying wave.

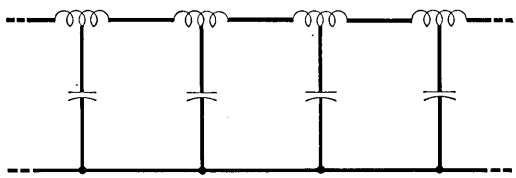

Fig. 5 — If the capacitances in this ladder network were negative it would support growing and decaying waves rather than traveling waves.

* Some care must be used in arriving at proper equivalent circuits. For instance, neither of the electric waves on a ladder network has negative energy if the network is set in motion, but we have seen that one of the longitudinal space-charge waves does have negative energy. If both the capacitances and the inductances of a ladder network are negative, the waves on the network will have negative energies.

become

$$\beta_1 = \frac{\omega}{u_0} + j\frac{\omega_q'}{u_0}$$

$$\beta_2 = \frac{\omega}{u_0} + j\frac{\omega_q'}{u_0}$$

The characteristic impedances become

$$K_1 = -j2\frac{\omega_q'}{\omega}\frac{V_0}{I_0}$$

$$K_2 = j2\frac{\omega_q'}{\omega}\frac{V_0}{I_0}$$

The fact that the characteristic impedances of the two waves are imaginary means that neither of the waves alone has any power flow. Neither of the waves can very well carry power. The amplitudes change with distance; hence for each wave Ui^* and iU^* increase or decrease with distance. But, the circuit and the electron beam are lossless, and the power cannot change with distance. As the waves do have a group velocity, neither has any stored energy. Does this mean that the beam cannot carry any power? The beam can carry power, just as a filter in its stop band can carry some power from a source at one end to a resistive load at the other end. The power flow is still given properly in terms of the total current i and the total voltage U by the same expression used in section 1. Suppose that the two waves have currents i_1 and i_2. Then the total power flow is

$$P = \tfrac{1}{2}[(i_1 + i_2)(K_1^* i_1^* + K_2^* i_2^*) + (i_1^* + i_2^*)(K_1 i_1 + K_2 i_2)]$$

$$P = \tfrac{1}{2}[(K_1 + K_1^*)(i_1 i_1^*) + (K_2 + K_2^*) i_2 i_2^* + i_1 i_2^*(K_1 + K_2^*) + (i_1 i_2^*(K_1 + K_2^*))^*]$$

First consider the case in which ω_q is real and for which the characteristis impedances are real and

$$K_1 = -K_2$$

In this case

$$P = K_1 i_1 i_1^* + K_2 i_2 i_2^*$$

This is the familiar case of unattenuated waves. The total power is the power of each wave calculated individually.

Let us now consider the case in which the effective plasma frequency

is imaginary. In this case we can write

$$K_1 = -jK_0$$

$$K_2 = +jK_0$$

where K_0 is real. We have

$$P = [-ji_1i_2^*K_0 + (-ji_1i_2^*K_0)^*]$$

Either wave alone carries no power; there is power flow only when the two waves are present simultaneously. The two waves vary with distance as

$$e^{-j(\omega/u_0)z}e^{(\omega_q'/u_0)z}; \qquad e^{-j(\omega/u_0/z)}e^{-(\omega_q'/u_0)z}$$

so the $i_1i_2^*$ is constant with distance. If this were not so the power would change with distance, but as the resonators have been assumed to be lossless, neither taking power from the beam nor adding power to the beam, this is impossible. Thus, in a lossless system an increasing wave is always one of a pair, and the other member decreases with distance in such a way as to keep the product of the amplitudes of the two waves constant with distance. Neither the increasing wave alone nor the decreasing wave alone carries any power, but the two together can carry power.

We will note that in the easitron the direction of the group velocity, that is, the direction of causality, is the direction of electron flow. Thus, the waves are both set up at the input resonator; it is there that boundary conditions on both current and voltage must be satisfied.

COUPLING OF MODES OF PROPAGATION

We know that waves which increase and decrease exponentially with distance are characteristic of a ladder network in which the susceptances of the shunt and series arms have the same signs. They occur in other networks as well. Consider a smooth transmission line loaded periodically with shunt capacitances, as shown in Fig. 6. Each capacitance reflects

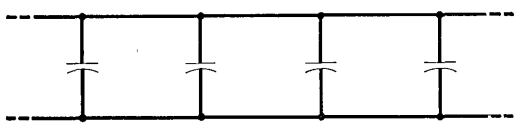

Fig. 6 — Capacitances connected across a smooth transmission line periodically couple the forward and backward waves and produce stop bands characterized by growing and decaying waves.

part of a wave approaching it. In other terms, each capacitance acts to couple one mode or wave (say the forward wave) to another (say, the backward wave). When the distance between the capacitances is such that the couplings reinforce, that is, near a half wavelength in this case, the system is a filter in its stop band; it does not transmit traveling waves, but supports rather a wave which increases exponentially with distance and a wave which decays exponentially with distance. Neither of these waves alone carries any power.

The space-charge waves of an electron stream can be coupled to one another, to a space-charge wave of another stream, or to an electromagnetic wave. In any of these cases we can have increasing waves.

THE SPACE-CHARGE-WAVE AMPLIFIER

Consider an electron beam surrounded by a series of metallic tubes A, B, A, B $\cdots$, alternately at different potentials with respect to the cathode from which the electrons come, as shown in Fig. 7. The impedances of the space-charge waves will be different in tubes A from what they are in tubes B. The behavior of this system is much like that for the transmission line system shown in Fig. 8, in which we have alternate line sections of different characteristic impedances K_A and K_B. We know that such a series of line sections forms a filter with stop bands.

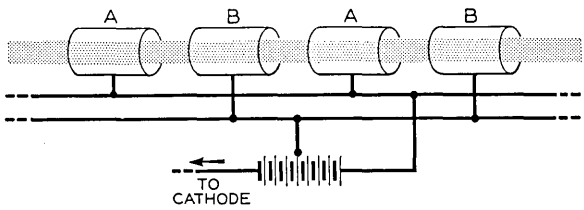

Fig. 7 — The impedances of waves in an electron beam passing through electrodes at alternately higher and lower potentials differ in regions of different potentials. This can result in stop bands characterized by growing and decaying waves. Such a device is a space-charge-wave amplifier.

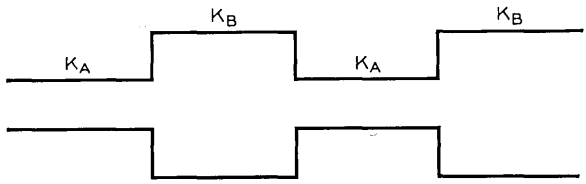

Fig. 8 — A transmission line with alternating sections of impedances K_A and K_B is somewhat analogous to the space-charge-wave amplifier.

In the case of the space-charge-wave structure of Fig. 7, the stop band occurs for conditions near that in which for both sections A and B the section lengths L_A and L_B are such that

$$2\frac{\omega_{qA}}{u_0}L_A = \pi$$

$$2\frac{\omega_{qB}}{u_0}L_B = \pi$$

Here ω_{qA} and ω_{qB} are the effective plasma frequencies for sections A and B.

A structure such as that of Fig. 7 can be interposed between input and output circuits, such as resonant cavities, to give a space-charge-wave amplifier dependent for its action on the growing wave of the pair.

THE TRAVELING-WAVE TUBE

In the space-charge-wave tube, the two waves which are coupled together have different velocities, just as the forward and backward waves on an electron stream have different velocities. Hence, they can be coupled strongly only through the use of some periodic structure in which the period is related to the difference in phase constants of the two waves.

In a traveling-wave tube we can have coupling between a space-charge wave and a wave traveling on a circuit, and both of the waves can have velocities which are nearly or exactly the same.

Here we must consider two different cases. If both of the coupled waves carry power in the same direction (that is, if the power is positive for both, or negative for both), coupling cannot result in a stop band, but only in transfer of power between one wave and the other. In order to have a stop band, power which we try to send in on one wave must come back to us on the other. Hence, to produce a stop band and gaining waves, the two coupled waves must carry powers with opposite signs.

A traveling-wave tube can consist of a helix of wire, which can support a slow electromagnetic wave, surrounding an electron beam, as shown in Fig. 9.

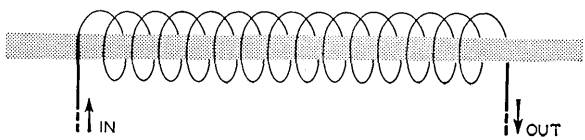

Fig. 9 — The vital elements of a traveling-wave amplifier are an electron stream and a slow-wave circuit which may be a helix surrounding the electron stream.

Traveling-wave tubes really involve at least four waves: two space-charge waves and two circuit waves. Usually, the backward circuit wave is so far out of synchronism with the space-charge waves that we can neglect its coupling with them. Further, if the space-charge waves are well separated in velocity, that is, when ω_q is large enough, then when one is coupled to the circuit wave the other isn't, and so we can get some idea of traveling-wave tube operation by considering waves in pairs. The simple mathematics of such coupling is given in Appendix D.

In Fig. 10, the behavior of various phase constants, plotted as a function of ω/u_0, is shown qualitatively. Here ω is radian frequency and u_0 is electron velocity. We may consider that ω/u_0 is varied by changing the electron velocity u_0 and keeping the frequency ω constant. The horizontal line β_c is the phase constant of the forward circuit wave in the absence of electrons, or when the coupling to the electrons is zero. β_c does not change with electron velocity. β_s and β_f are the phase constants of the slow and fast space-charge waves, respectively, with zero coupling to the circuit wave. For the slow space-charge wave, the power flow is negative, while for the circuit wave and the fast space-charge wave the power flow is positive. Thus, for coupling between the slow

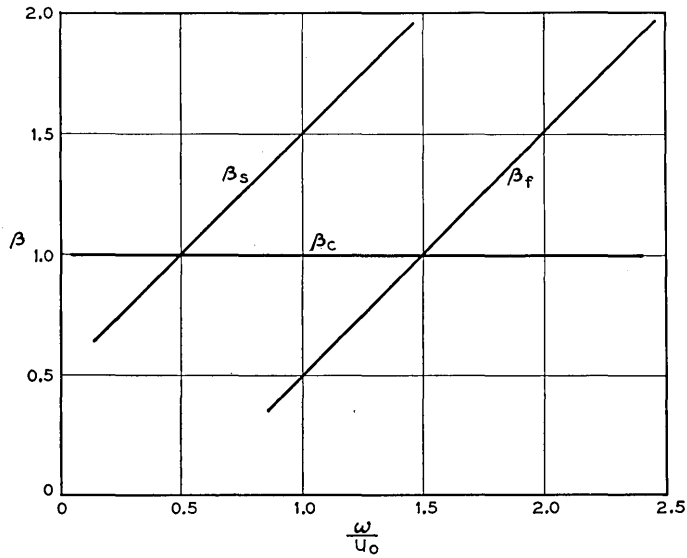

Fig. 10 — Suppose that at a constant radian frequency ω we change the electron velocity u_0 in a traveling-wave tube. If the waves of the electron stream were not coupled to the waves of the helix, the phase constants, β_c of the forward circuit wave, β_s of the slow wave, and β_f of the fast wave, would vary approximately as shown.

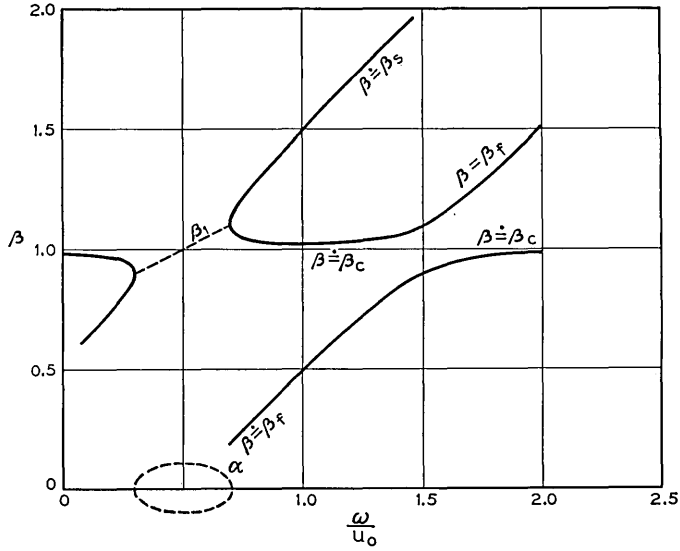

Fig. 11 — Because of coupling of the space-charge waves to the forward circuit wave, gain is produced near $\beta_c = \beta_s$, while the curves sheer off from one another near $\beta_c = \beta_f$.

space-charge wave and the circuit wave we can have a stop band, while for coupling between the circuit wave and the fast space-charge wave we cannot.

The consequences of the couplings between the circuit wave and the space-charge waves near the intersections of β_c with β_s and β_f are illustrated in Fig. 11.

We see from Fig. 11 that near synchronism between the circuit wave and the fast space-charge wave ($\beta_c = \beta_f$ for no coupling) these waves combine so that for any given value of ω/u_0 there are always two distinct real values of β. This is typical for coupling between modes with power flows of the same sign. At $\beta_c = \beta_f$ each of the two mixed waves has equal energies in the circuit and in the electron stream.

Near synchronism between the circuit wave and the slow space-charge wave ($\beta_c = \beta_s$ in absence of coupling) these two waves combine so that over a range of ω/u_0 near $\beta_c = \beta_s$, β has two complex values with the same real part and with equal and opposite imaginary parts. We can write this as

$$\beta = \beta_1 \pm j\alpha; \qquad -j\beta = -j\beta_1 \mp \alpha$$

This corresponds to an attenuated and a growing wave with the same phase velocity. In Fig. 11, β_1 and α are plotted as dashed lines.

Over the range of ω/u_0 for which the waves are attenuated ($\alpha \neq 0$) the net power flow in each of the modes is zero. The power flow in the electron stream is equal and opposite to that in the circuit. Such behavior is characteristic when two modes with power flows of opposite signs are coupled. It is characteristic of the stop band of an electric wave filter.

The curves of Fig. 11 exhibit the same behavior that has been found by other means, although similar curves are sometimes plotted somewhat differently.

When an input signal is applied to the helix of a traveling-wave tube, all three forward waves are set up. The increasing wave grows until it predominates, and it forms the amplified output of the tube.

The total ac power of the increasing wave is zero. How can we obtain power from it? In the increasing wave we have a positive electromagnetic power flow in the circuit and an equal negative power flow in the electron stream. If we terminate the helix we can draw off the electromagnetic power; the electron stream is left with less power than it had on entering the helix.

DOUBLE-STREAM AMPLIFIERS

A double-stream amplifier makes use of two streams of electrons which have different velocities, as shown in Fig. 12. The behavior of a double-stream amplifier is very similar to that of a traveling-wave tube. In such a device each electron stream supports a slow, negative-energy wave and a fast, positive-energy wave. At a constant frequency ω let the velocity u_1 of one stream be kept constant and let the velocity u_2 of the other stream be varied. The behavior of the phase constants β of the waves is shown qualitatively in Fig. 13. β_{s1} and β_{f1} are the phase constants of the slow and fast waves of the constant-velocity stream, and β_{s2} and β_{f2} are the phase constants of the slow and fast waves of the stream whose velocity is changed. There are two ranges of velocity u_2 for which gain is obtained; for u_2 a little larger or a little smaller than u_1.

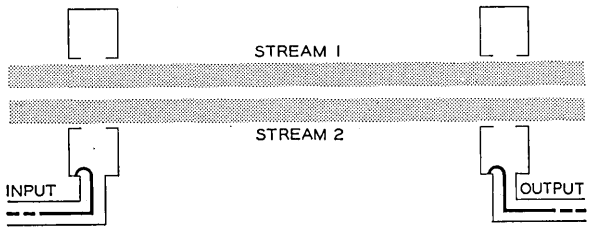

Fig. 12 — Two nearby electron streams of different velocities u_1 and u_2 constitute a double-stream amplifier.

NOISE WAVES ON ELECTRON STREAMS

Consider the electrons of the beam as they leave the cathode. If the velocity distribution is Maxwellian, and if the electrons leave independently, there will be a mean-square fluctuation in convection current, i^2, given by

$$i^2 = 2eI_0B$$

and an uncorrelated mean square fluctuation in ac velocity, v^2, given by

$$v^2 = (4 - \pi)\left(\frac{e}{I_0}\right)\left(\frac{kT_c}{m}\right)B$$

Here I_0 is beam current, e and m are electron charge and electron mass' k is Boltzmann's constant, T_c is cathode temperature and B is bandwidth.

Usually, space-charge-limited flow is used. In this case the beam current is only a part of the emitted current; the rest is turned back at the potential minimum. In this case we may use the above relations, counting I_0 as the beam current, as some sort of approximation for the current passing the potential minimum.

The wave picture we have been discussing may be seriously inaccurate near the cathode where the relative spread in electron velocities is large. Suppose that we hope for the best and apply it. We find that in the most general case our electron stream will have on it a noise standing-wave pattern. If $i_{\min}$ and $i_{\max}$ are the minimum and the maximum noise currents,

$$\frac{|\, i_{\min}\,|\,|\, i_{\max}\,|}{2eI_0B} = \frac{1}{2}\,\alpha\left(\frac{\omega}{\omega_q}\right)\left(\frac{kT_c}{eV_0}\right)$$

Here α is a constant near to unity.

The noise pattern is made up of two uncorrelated noise standing-wave patterns, one from i at the cathode and the other from v at the cathode; these patterns have amplitudes i_1 and i_2 at their maxima; the minima are of course zero. We have

$$|\, i_{\min}\,|\,|\, i_{\max}\,| = |\, i_1\,|\,|\, i_2\,|\sin\Psi$$

Here Ψ is the relative phase angle of the standing-wave patterns associated with i_1 and i_2. That is, if the maximum of the i_2 pattern is at that of i_1, $\Psi = 0$, while if the maximum of the i_2 pattern is midway between maxima of i_1, then $\Psi = \pi/2$.

The first of these theorems says something about the noise current at the maximum and that at the minimum, but it does not directly say how

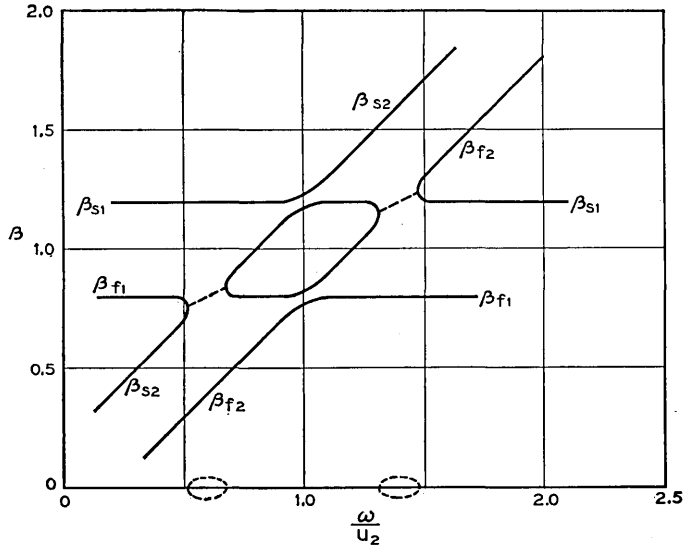

Fig. 13 — In a double-stream amplifier, gain is obtained when the phase constants of the slow wave of the faster stream and the fast wave of the slower stream are nearly equal.

large the maximum is. For an ordinary two-potential electron gun, i_{max} is very large compared with i_{min} .

NOISE DEAMPLIFICATION

Early traveling-wave tubes made use of a two-potential electron gun spaced a critical distance from the circuit, as shown in Fig. 14. More recently it has been found possible to reduce the noise figure considerably by the use of space-charge-wave amplification, as discussed in the section on "The Space-Charge-Wave Amplifier." The structure used is indicated in Fig. 15. The gun has a low-potential anode followed by a

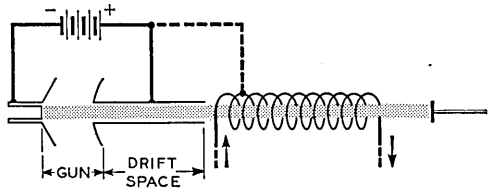

Fig. 14 — When a simple, two-potential electron gun is used, the noise figure of a traveling-wave tube can be optimized by adjusting the drift-space between the gun anode and the helix.

drift tube. At the point where the noise current is a minimum the voltage is "jumped" to the helix voltage. A second drift tube follows, so that there is a critical distance between the jump and the helix.

The effect of this "voltage jump" gun is to deamplify the component of the space-charge waves which is associated with the noise current at the current maximum. In space-charge-wave amplifier terms, this component sets up the decreasing wave only. Thus, in the second drift tube the ratio $| i_{max}/i_{min} |$ is smaller than in the first.

By using a single velocity jump, traveling-wave tubes with noise figures around 8 db have been made.

The use of more velocity jumps has been proposed. It can be shown, however, that as i_{max} is deamplified, i_{min} must be amplified. This sets a theoretical limit of around 6 db to the noise figure attainable by means of space-charge-wave deamplification alone.

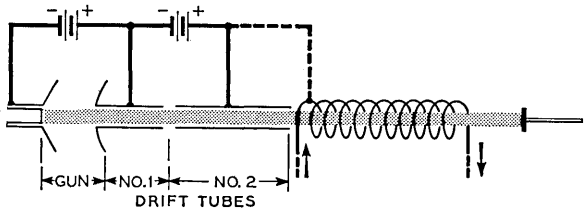

Fig. 15 — When a two-potential or "velocity jump" gun is used, the noise figure can be reduced by space-charge-wave deamplification of the noise on the electron stream.

NOISE CANCELLATION

It would be highly desirable to build a traveling-wave tube such that the electromagnetic input would excite an increasing wave, but the noise in the electron stream would excite only some combination of the decreasing and the unattenuated waves. If we succeeded in this, the noise introduced by the tube could be made as small relative to the signal as desired, merely by making the tube long enough. Can we accomplish this by means of some special structure near the input end of the tube?

We can represent the noise on the electron stream at some reference point by means of a velocity fluctuation v and a current fluctuation i; we have seen that neither can be zero. Because the system is linear, superposition applies, and the amplitudes of the growing, attenuated and unattenuated waves which are excited are the sums of the amplitudes excited by i and v independently.

Suppose that $v = 0$. Then the beam carries no power. Thus, i cannot

excite the unattenuated wave, for that wave carries power. Let us assume that it excites the decreasing wave alone, which, when present alone, carries no power. So far there is no contradiction, and we can believe that it is possible to arrange matters so that the current alone does not excite the increasing wave.

Suppose we have so arranged matters that i does excite the decreasing wave only. Consider what happens when $i = 0$. Can v alone excite the decreasing wave only if i alone excites the decreasing wave only? If it can, then v and i together must excite the decreasing wave only. But suppose v and i are of the same frequency and in phase. Then the beam carries power. But, the decreasing wave alone cannot carry power, and hence what we have assumed is impossible. If the i excites the decreasing wave only, then v must excite at least a component of the growing wave. Hence, we cannot cancel out the noise from the beam completely.

FINAL COMMENTS

We have seen that the properties of space-charge waves and the behavior which must follow when space-charge waves are coupled to other space-charge waves or to circuit waves can be used to explain the operation of seemingly diverse types of microwave tubes. The wave picture gives a clear and quantitative picture of energy relations and power flow. It enables us to understand simply the effect of thermal velocities on the operation of tubes through their effect on the phase constants of the space-charge waves. It is useful in detailed considerations of noise, and in one case it has enabled us to draw a general conclusion without resorting to formal mathematical manipulation. It may well be that the wave picture can be of further use both in calculating detailed behavior of tubes and in understanding their general properties.

Appendix A

SPACE-CHARGE WAVES

Consider a narrow electron stream in which we may assume that electron velocity and charge density do not vary across the stream, and in which the electrons are free to move in the z-direction only. An axially symmetrical electron focusing system immersed, cathode and all, in a strong magnetic field approximates this.

Let all ac quantities contain the factor

$$e^{-j\beta z}e^{j\omega t}$$

and let the total charge density, current and electron velocity be made

up of dc and ac parts as follows:*

charge density: $-\rho_0 + \rho$

convection current density: $-I_0 + i$

velocity: $u_0 + v$

Here ρ_0, I_0 and u_0 are positive dc quantities. The quantities on the right are the ac components.

We have from the definition of convection current

$$(-I_0 + i) = (-\rho_0 + \rho)(u_0 + v) \tag{A1}$$

In the case of very low level operation, we neglect products of ac quantities in comparison with products of ac and dc quantities. Doing this, we obtain from (A1) the dc and ac convection currents

$$I_0 = \rho_0 u_0 \tag{A2}$$

or

$$i = -\rho_0 v + u_0 \rho \tag{A3}$$

$$\rho = \frac{i + \rho_0 v}{u_0} \tag{A4}$$

We can apply the continuity equation, or, the equation of conservation of charge, to the ac convection current

$$\frac{\partial i}{\partial z} = -\frac{\partial \rho}{\partial t} \tag{A5}$$

$$\beta i = \omega \rho$$

$$\beta i = \left(\frac{1}{u_0}\right)(j\omega i - \rho_0 j\omega v)$$

$$i = \frac{-\omega \rho_0 v}{\omega - \beta u_0} \tag{A6}$$

Thus, if we have a wave with a given phase constant β, and if we know ρ_0 and ω, (A6) gives the convection current in terms of ac electron velocity. How can we find what β will be? To find this we must consider the effect of the electric field on the electrons. Consider an ac electric field E_z in the z direction, which also varies with time and distance as

* It will be convenient elsewhere to use $-I_0$ and i as currents rather than current densities and $-\rho_0$ and ρ as charge per unit length.

do the other ac quantities. We can write

$$\frac{dv}{dt} = -\frac{e}{m} E_z \tag{A7}$$

Here e/m, the charge-to-mass ratio of the electron, is taken as a positive quantity.

In (A7), dv/dt is the rate of change of v with respect to t for a single electron; that is dv/dt observed riding along with the electron. If we ride along with the electron for a time dt we move along distance dz

$$dz = (u_0 + v)\, dt$$

For small signals we neglect v in this expression and write

$$dz = u_0\, dt$$

Hence, the total change dv in the velocity of the electron in the time dt is

$$dv = \frac{\partial v}{\partial t}\, dt + \frac{\partial v}{\partial z}\, u_0\, dt$$

Hence, we find that

$$\frac{dv}{dt} = j(\omega - \beta u_0)v \tag{A8}$$

Using (A7) and (A8), we see that

$$v = \frac{j\,\frac{e}{m} E_z}{(\omega - \beta u_0)} \tag{A9}$$

We can combine (A9) with (A6) and write for the convection current density

$$i = \frac{-j\,\frac{e}{m}\,\omega\rho_0 E_z}{(\omega - \beta u_0)^2} \tag{A10}$$

Let us now consider a special, hypothetical case in which the electric field is in the z direction only, so that there are no transverse electric

fields and no transverse displacement current. Then the total ac current density i_t is the sum of the convection current density and the displacement current density, or,

$$i_t = i + j\omega\epsilon E_z$$

$$i_t = \left(\frac{-\frac{e}{m}\rho_0}{\epsilon(\omega - \beta u_0)^2} + 1 \right) j\omega\epsilon E_z \tag{A11}$$

Let us use a quantity ω_p, which was long ago named the plasma frequency (radian frequency)

$$\omega_p^2 = \frac{\frac{e}{m}\rho_0}{\epsilon} \tag{A12}$$

Using ω_p, (A10) can be written as

$$i_t = \left(-\frac{\omega_p^2}{(\omega - \beta u_0)^2} + 1 \right) j\omega\epsilon E_z \tag{A13}$$

According to Maxwell's equations the divergence of the total current is zero. Both components of i_t vary with z. If, as we have assumed, there is no current normal to the z direction, then i_t must be zero. If this is to be so, we must have

$$(\omega - \beta u_0)^2 = \omega_p^2$$

$$\beta = \frac{\omega}{u_0} \pm \frac{\omega_p}{u_0} \tag{A14}$$

In actual electron beams there is transverse electric field away from the beam, and hence i_t is not zero. It is found, however, that when ω_p is small compared with ω, we can write quite accurately

$$\beta = \frac{\omega}{u_0} \pm \frac{\omega_q}{u_0} \tag{A15}$$

Here ω_q, which is known as the *effective plasma frequency*, is smaller than ω_p. As ω is raised, so that the wavelength of the space-charge waves becomes smaller compared with the diameter of the electron beam, the electric field tends to become largely longitudinal and ω_q approaches ω_p.

The upper sign in (A15) gives the phase constant of the slow wave, a wave with a phase velocity less than that of the electrons. The lower sign gives that of the fast wave, a wave with a phase velocity faster than that of the electrons.

From (A15) and (A6) we note that

$$i = \pm\frac{\omega}{\omega_q}\rho_0 v \tag{A16}$$

The upper sign holds for the slow wave; the lower sign for the fast wave.

It has been convenient to use $-I_0$ and i as current densities and $-\rho_0$ and ρ as charge densities. In subsequent work and in the text, $-I_0$ and i will be used as beam current and $-\rho_0$ and ρ as charge per unit length. All the relations of this appendix except (A11)–(A13) will hold if the quantities are so interpreted.

Appendix B

Power Flow in Space Charge Waves

The purpose of this appendix is to justify the expression for power flow in the beam.

Consider that the electron beam is acted on over a short distance by an ac voltage. Imagine, for instance, that the beam passes through two very closely spaced grids which form a part of a resonator, and that a voltage ΔV appears between the grids. What does the voltage do to the beam?

The voltage ΔV changes the velocity of the electrons but it does not change the convection current. To find out how much the velocity is changed we need only consider the case in which the beam has no ac velocity on reaching the grids, since in a linear system the change will be the same in all other cases. The total velocity $u_0 + v$ is given in terms of the total accelerating voltage $V + \Delta V$ by

$$u_0 + v = \sqrt{2\frac{e}{m}(V_0 + \Delta V)} \tag{B1}$$

We assume ΔV to be small, so that

$$v = \frac{\Delta V \frac{e}{m}}{\sqrt{2\frac{e}{m}V_0}} = \frac{\Delta V \frac{e}{m}}{u_0} \tag{B2}$$

The change ΔU in the "voltage" U is

$$\Delta U = -\frac{u_0 v}{\frac{e}{m}} = -\Delta V \tag{B3}$$

The convection current i flows against the voltage ΔV, so that a power ΔP is transferred from the beam to the resonator which is attached to the grids.

$$\Delta P = -Re\Delta V i^* \tag{B4}$$

Thus, the change in the power in the beam in passing through the grids must be $-\Delta P$

$$-\Delta P = Re\,(-\Delta V i^*) = Re\Delta U i^* \tag{B5}$$

$$\Delta P = -Re\Delta U i^*$$

According to the expression we have used in calculating beam power, if the "voltage" of the beam on reaching the grids is U, and the convection current is i, then the beam power P_1 on reaching the grids is

$$P_1 = ReUi^* \tag{B6}$$

After passing through the grids, U is increased by an amount ΔU while the current is unchanged, so that the power P_2 of the beam leaving the grids is

$$P_2 = Re(U + \Delta U)i^* \tag{B7}$$

The loss of power in the beam, ΔP, is

$$\Delta P = P_1 - P_2 = -Re\Delta U i^* \tag{B8}$$

This agrees with (B5), in which ΔP was calculated as the power lost from the beam to the resonator.

Appendix C

The Effect of the Velocity Distribution in the Electron Beam on the Effective Plasma Frequency

Consider an electron beam in which electron motion is confined to the z-direction, and in which the electrons have a velocity spread with a

mean square deviation $\langle u^2 \rangle$ about the mean value u_0. If ω_{q0} is the effective plasma frequency for $\langle u^2 \rangle = 0$, then taking the velocity spread into account, the effective plasma frequency ω_q is given approximately by

$$\omega_q^2 = \omega_{q0}^2 \left(1 + 3 \frac{\langle u^2 \rangle}{u_0^2} \left(\frac{\omega}{\omega_{q0}}\right)^2\right)$$

If the velocity distribution is the same as for electrons accelerated individually by a voltage V_0, that is, if electron interactions do not affect the velocity distribution appreciably (as they probably do not)

$$\frac{\langle u^2 \rangle}{u_0^2} = \frac{1}{4}\left(\frac{kT_c}{eV_0}\right)^2 = \frac{1}{4}\left(\frac{T_c}{11{,}600\ V_0}\right)^2$$

Here k is Boltzman's constant and T_c is cathode temperature. Thus, from this assumption

$$\omega_q^2 = \omega_{q0}^2 \left(1 + \frac{3}{4}\left(\frac{kT_c}{eV_0}\right)^2 \left(\frac{\omega}{\omega_{q0}}\right)^2\right)$$

Following our wave picture, we can take into account the thermal velocity spread by using this corrected value for the effective plasma frequency in all our formulae. For all practical purposes, the change in effective plasma frequency due to thermal velocities is negligible.

In a paper which will appear in the Journal of Applied Physics, D. A. Watkins has used a somewhat different approach in treating the effect of thermal velocities on the operation of traveling-wave tubes.

Appendix D

Phase and Attenuation Curves for Coupled Modes

When two unattenuated modes of propagation are coupled together periodically in a lossless manner, they combine to form two new modes. For each of these new modes the amplitude is changed in one period of the coupling structure by a factor

$$Me^{-j(\theta_q - \theta_p)/2} e^{j(\theta_3 + \theta_1)/2} \tag{D1}$$

where M is a root of

$$M^2 - 2\sqrt{1 \mp k^2} \cos \frac{(\theta_q - \theta_p + \theta_1 - \theta_3)}{2} + 1 = 0 \tag{D2}$$

Here k is a coupling coefficient which is zero for zero coupling. The upper sign applies if the power flow in the two modes have the same

signs while the lower sign applies if the power flows have opposite signs. θ_q and θ_p are phase lags per coupling period associated with the two original modes and θ_1 and θ_2 are phase angles associated with the coupling device.

We can treat the case of continuous coupling by letting the period of coupling L be very short, the angles θ_q , θ_p , θ_1 , θ_3 be very small, and the coupling per period, k, be very small. In this case the cosine can be represented by the first terms of a power series and we find that the phase constants β of the modes are given by

$$\beta = \frac{\beta_a + \beta_b}{2} \pm \left(\frac{\beta_a - \beta_b}{2}\right)\sqrt{1 \pm \left(\frac{2K}{\beta_a - \beta_b}\right)^2} \tag{D3}$$

Here β_a and β_b are the phase constants for $K = 0$ (zero coupling)

$$\beta_a = \frac{\theta_p - \theta_1}{L} \tag{D4}$$

$$\beta_b = \frac{\theta_q - \theta_3}{L} \tag{D5}$$

and K is the coupling per unit length

$$K = \frac{k}{L} \tag{D6}$$

As before, the upper sign in the radical applies when the power flows have the same signs and the lower sign when the power flows have opposite signs.

In applying (D3) to the case of traveling-wave tubes and backward-wave oscillators, the effect of all but two modes was of course neglected when the two phase constants would have had nearly the same value in the absence of coupling; the curves for such regions were then joined smoothly to give the overall plots of Figs. 11 and 13.

In Fig. 11 the parameters chosen arbitrarily were:

$$\beta_c = 1$$

$$\beta_s = \omega/u_0 + \tfrac{1}{2}$$

$$\beta_f = \omega/u_0 - \tfrac{1}{2}$$

$$K = 0.1$$

The complex portion of the phase constant, or, the real portion of the propagation constant, in a stop band caused by the coupling of two

modes with power flows of opposite signs is designated by α and is plotted as the dashed ellipses about the horizontal axis.

Appendix E

This appendix comments briefly on various sections and cites references, which are listed at the end of the appendix. The list of references is not exhaustive, but it should enable the interested reader to follow work back to its source.

1. Space-Charge Waves

Space-charge waves of the general sort considered are related to the plasma oscillations of Tonks and Langmuir.[1] Waves in long beams were first discussed by Hahn[2] and Ramo.[3] The effects of a velocity distribution are discussed by Pierce[4] and by Bohm and Gross.[5] The negative energy of the slow space-charge wave has been reported by Chu[6] and by Walker.[7] Chu gave the effective "voltage" U and the characteristic impedance K for the waves.

2. The Klystron

Beck[8] gives an adequate description of and references to klystrons.

3. The Resistive Wall Amplifier

The effect has been discussed by Pierce,[9] and a tube using it has been described by Birdsall, Brewer and Haeff.[10]

4. The Easitron; Increasing Wave in a Lossless System

The original easitron was a tube built by L. R. Walker at Bell Telephone Laboratories; it was a 3-cm tube using half-wave wires as resonant elements. It has not been described in the literature. Pierce has discussed the operation of this sort of multi-resonator klystron on page 195 of *Traveling Wave Tubes*[11] and elsewhere.[9]

5. Coupling of Modes of Propagation

The operation of traveling-wave tubes was first explained in terms of coupling between an electromagnetic wave and a space-charge wave by C. C. Cutler in unpublished work. Mathews has made an analysis in

these terms.[13] Such coupling has been considered in general terms by Pierce.[12]

6. The Space-Charge-Wave Amplifier

This tube was invented by Tien, Field and Watkins[14] and is described in more detail by Tien and Field.[15]

7. The Traveling Wave Tube

Adequate descriptions and references are available in work by Kompfner,[16] Pierce,[11] and Beck.[8]

8. Double-Stream Amplifiers

Descriptions and references are given by Pierce[11] and by Beck.[8]

9. Noise Waves in Electron Streams

Cutler and Quate have published experimental results.[17] The theorems quoted are given by Pierce.[18]

10. Noise Deamplification

This was suggested by Tien, Field and Watkins[14] and is described in detail by Watkins[19] and Peter.[20]

11. Noise Cancellation

Noise cancellation was first proposed by C. F. Quate.[21]

REFERENCES

1. Lewi Tonks and Irving Langmuir, Phys. Rev., **33,** pp. 195–210 and p. 990, 1929.
2. W. C. Hahn, Gen. Elect. Rev., **42,** pp. 258–270, 1939.
3. Simon Ramo, Phys. Rev., **56,** pp. 276–283, 1939.
4. J. R. Pierce, Jour. App. Phys., **19,** pp. 231–236, 1948.
5. D. Bohm and E. P. Gross, Phys. Rev., **75,** pp. 1851–1876, 1949, **79,** pp. 992–1001, 1950.
6. L. J. Chu, paper presented at the Institute of Radio Engineers Electron Devices Conference, University of New Hampshire, June, 1951.
7. L. R. Walker, J. App. Phys., **25,** pp. 615–618, May, 1954.
8. *Thermionic Valves,* A. H. W. Beck, Cambridge University Press, 1953.
9. J. R. Pierce, B.S.T.J., **30,** pp. 626–651, 1951.
10. Charles K. Birdsall, George R. Brewer and Andrew V. Haeff, Proc. I. R. E., pp. 865–871, 1953.
11. *Traveling Wave Tubes,* J. R. Pierce, Van Nostrand (1950).
12. W. E. Mathews, J. App. Phys., **22,** pp. 310–316, 1951.

13. J. R. Pierce, J. App. Phys., **25,** pp. 179–183, Feb. 1954.
14. Ping King Tien, Lester M. Field and D. A. Watkins, Proc. I.R.E., **39,** p. 194, 1951.
15. Ping King Tien and Lester M. Field, Proc. I.R.E., **40,** pp. 688–695, 1952.
16. R. Kompfner, Rep. Progress. Phys., **15,** pp. 275–327, 1952.
17. C. C. Cutler and C. F. Quate, Phys. Rev., **80,** pp. 875–878, 1950.
18. J. R. Pierce, J. App. Phys., **8,** pp. 93–933, 1954.
19. D. A. Watkins, Proc. I.R.E., **40,** pp. 65–70, 1952.
20. R. W. Peter, R.C.A. Review, **13,** pp. 344–368, 1952.
21. C. F. Quate, paper presented at the Institute of Radio Engineers Electron Devices Conference, University of New Hampshire, 1952.

Theory of Open-Contact Performance of Twin Contacts

By M. M. ATALLA and MISS R. E. COX

(Manuscript Received June 17, 1954)

The first part is a presentation of an analytical study of the open-contact performance of twin contacts. It provides means for predicting their performance from single contact data. It is shown that the probability of failure of twin contacts is generally appreciably greater than the square of the probability of failure of single contacts. This is supplemented with the results of an experimental study which determines the effects of a few design parameters on the performance of single contacts. These are the parameters that determine the magnitude of improvement in performance obtained by replacing single contacts by twin contacts.

INTRODUCTION

The present switching apparatus normally operates in atmospheres that may be contaminated with dust particles and foreign matter. Some apparatus components, particularly the contacts, are relatively sensitive to such contaminations which may interrupt the proper functioning of a pair of contacts. Normally, a single switching operation in a central office requires the operation of as many as a thousand relays or 10,000 contacts. To secure the high level of performance desired, it is evident that superlative performance and high degree of reliability of the contacts are essential.

Many attempts have been and are being made to reduce the so-called "open" contact troubles due to foreign matter. Examples of environmental precautions are filtering the air supply to the central office, enclosing apparatus in cabinets, limiting personnel activities in the office, etc. An additional precaution incorporated in the apparatus design is the use of twin contacts. Such a scheme, when *properly* used, should result in substantial improvement in performance since an open can only take place when both members become open simultaneously. It may occur to one that the probability of a twin-contact open is the square of the prob-

ability of a single-contact open. Such a performance, however, has never been observed in practice where an improvement of only 10:1 is usually more typical. A study of the mechanisms involved has revealed that only a very small number of opens are obtained due to the simultaneous occurrence of opens on both members of a twin contact. The majority of opens, however, occur by having an open in one member of a twin contact which persists long enough to allow the occurrence of an open on the other member. By expressing this physical process in mathematical terms it was possible to develop a theory of performance of twin contacts in terms of the characteristics of single contacts.

NOTATION

- d Diameter of dust particle
- f Fractions of opens in single contacts cleared after N operations
- f_∞ The asymptotic value of f corresponding to $N = \infty$
- $\bar{n}$ Average number of operations required to clear an open on a single contact
- r Distance of particle from center of circular open contact zone
- r_0 Radius of "open zone"
- s Fraction of the twin contacts that are half open at any time, $s = s_\infty + s_{\bar{n}}$
- s_∞ Fraction of twin contacts that are permanently half open
- $s_{\bar{n}}$ Fraction of twin contacts that are temporarily half open
- w Mechanical wipe
- x Average displacement of a dust particle per contact operation
- F Contact force
- N Number of contact operations
- P_s Probability of occurrence of a single-contact open in opens/contact operation
- P_t Probability of occurrence of a twin-contact open in opens/contact operation
- X Total displacement distance to clear an open
- α $= (2 - P_s)(1 - f_\infty)$
- β $= \bar{n} f_\infty P_s(2 - P_s)$
- θ Angle of displacement
- φ Slope of contact surface irregularity

PRESENTATION OF THEORY

Outline and Assumptions

Consider a large group of twin contacts each constituting a pair of identical and entirely independent single contacts. After a period of

operation a certain number of the twin contacts will become half-open.* These contacts will behave as if they were single-contacts until either: (1) the open half clears itself by operation, or (2) the other half becomes open leading to a twin-contact open. It is assumed that a failing twin-contact is cleared by an operator and then put back into service.

In developing the theory it was necessary to represent by analytic expressions the rate of occurrence of opens on single contacts and the rate of their clearing by operation. These are approximations of a fairly large amount of experimental data consistently obtained from a number of tests on a variety of actual telephone relay contacts.

(1) Rate of opens of single contacts "P_s": For a large number of single contacts at one set of operating conditions, the rate of opens is usually constant. This constant depends primarily on the quality and concentration of the offending foreign matter involved, the design of the contacts and their mechanism of actuation. Fig. 1 shows the results of three tests on single contacts of different design at different test conditions. They all substantiate the assumption that the rate of opens of

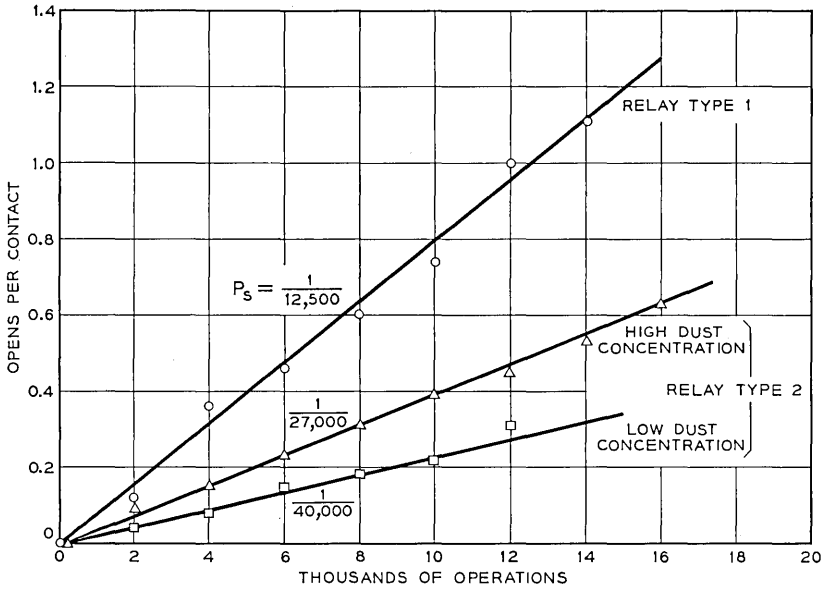

Fig. 1 — Rate of opens of single relay contacts.

* A half-open is defined as one where only *one* member, of a pair in a twin-contact, is open. In practice, a half-open is not normally detected. Only a simultaneous open on both members of a twin contact will cause a circuit failure.

single contacts is a constant:

$$P_s = \text{constant} \tag{1}$$

where P_s is defined as the number of opens per "contact operation." In some cases, a certain transient period may precede the equilibrium characteristic "P_s = constant characteristic." This transient period is usually relatively short and is neglected in this analysis.

(2) Clearing of opens on single-contacts by operation: If an open single contact is allowed to operate mechanically, it is possible that it will clear itself after a number of operations. As discussed in a later section, the opens obtained are never identical in nature. They, instead, have a certain statistical distribution which usually accounts for a wide spread in their clearing rate. If, however, the operating conditions are under control, the clearing characteristic of a set of contacts is found to follow a well defined and reproducible statistical distribution. Fig. 2 shows an accumulative distribution curve for clearing opens produced by cotton lint fibres.* The ordinate represents the fraction of the open single contacts that clear after N operations as given by the abscissa. In general, these relations have the following typical characteristics. The first operation following the occurrence of the open is the most efficient† single operation in clearing opens. It is usually responsible for clearing 10 to 30 per cent of the total number of opens. The subsequent operations are progressively less efficient and in general a certain fraction

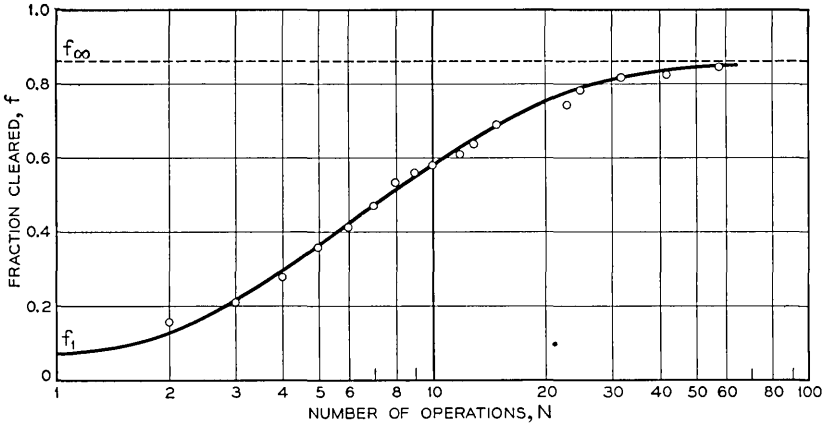

Fig. 2 — Distribution of clearing opens caused by lint.

* This is one of the major causes of open contacts in central offices. These fibres are usually in ribbon form of various configurations.

† This apparent efficiency is only due to the presence of opens that are more easy to clear than others. These will readily clear after one or a few operations.

$(1 - f_\infty)$ of the opens will persist for a relatively large number of operations. A study of a variety of these clearing characteristics generally indicates a rapid rise to the asymptotic value f_∞ in less than 100 operations, and to $f = 0.5$ in less than 10 operations. As will be shown the fractional persistency $(1 - f_\infty)$ is of major importance in determining twin-contact performance. In general, however, opens on twin-contacts are due to both the persistent half-opens and the temporary half-opens that might develop into twin-opens before clearing takes place.

DEVELOPMENT OF THE THEORY

Consider a large number of contacts operating at steady conditions. After N operations, let s_∞ be the fraction of the contacts that is permanently half-open and $s_{\bar{n}}$ be the fraction that is temporarily half-open. As discussed, the number of operations necessary to clear a half-open is not constant and, for the majority of the contacts, is of the order of a few operations. To simplify the treatment, it is assumed that each temporary half-open will clear in an average of $\bar{n}$ operations *from the time it first occurred.* The fraction $s_{\bar{n}}$ must, therefore, have been produced during the $\bar{n}$ operations directly preceding the time t. Since $\bar{n}$ is usually relatively small, the universe can be assumed to have had a negligible change during the operations $\bar{n}$. Hence,

$$\begin{aligned} s_{\bar{n}} &= (\text{rate of formation of temporary half-opens}) \times \bar{n} \\ &= \bar{n}[2(1 - s)P_s - (1 - s)P_s^2]f_\infty \end{aligned}$$

where s = total fraction of half-opens $= s_{\bar{n}} + s_\infty$. Substituting $\beta = \bar{n}f_\infty P_s(2 - P_s)$ one gets

$$s_{\bar{n}} = \frac{\beta}{1 + \beta}(1 - s_\infty) \tag{2}$$

Also, after N operations, the incremental change ds_∞ due to dN operation is:

$$ds_\infty = [2(1 - s)P_s - (1 - s)P_s^2](1 - f_\infty)\,dN - s_\infty P_s\,dN$$

where the second term is the reduction in s_∞ due to occurrence of twin-opens. Substituting $\alpha = (2 - P_s)(1 - f_\infty)$ and combining with equation 2 to eliminate $s_{\bar{n}}$ give:

$$ds_\infty = -\frac{1 + \alpha + \beta}{1 + \beta}\left[s_\infty - \frac{\alpha}{1 + \alpha + \beta}\right]P_s\,dN \tag{3}$$

$$s_\infty = \frac{1}{1 + \alpha + \beta} + Ke^{-(1+\alpha+\beta)/(1+\beta)P_s N} \tag{4}$$

where K is an integration constant. Let at $N = 0$, $s_\infty = s_0$, which is a description for the initial conditions of the contacts. The solution becomes:

$$s_\infty = \frac{\alpha}{1+\alpha+\beta}\left[1 - \left(1 - \frac{1+\alpha+\beta}{\alpha}\, s_0\right)e^{-(1+\alpha+\beta)/(1+\beta)P_s N}\right] \tag{5}$$

The rate, or probability, of twin-contact failure is determined from:

$$P_t = sP_s + (1 - s)P_s^2 \tag{6}$$

Substituting from 2 and 5 into 6 and reducing gives:

$$P_t = P_s^2 + P_s(1 - P_s)\,\frac{\beta}{1+\beta}$$
$$\left[1 + \frac{\alpha/\beta}{1+\alpha+\beta}\left(1 - \left(1 - \frac{1+\alpha+\beta}{\alpha}\, s_0\right)e^{-(1+\alpha+\beta)/(1+\beta)P_s N}\right)\right] \tag{7}$$

where, one may repeat for convenience:

$$\alpha = (2 - P_s)(1 - f_\infty) \text{ and } \beta = \bar{n} f_\infty P_s(2 - P_s)$$

For all practical cases, $P_s \ll 1$, $\alpha = 2(1 - f_\infty)$ and $\beta = 2\bar{n} f_\infty P_s < 1$. Substituting in 7 gives

$$P_t = P_s^2 + 2P_s$$
$$\left[\bar{n} f_\infty P_s + \frac{1 - f_\infty}{3 - 2f_\infty}\left(1 - \left(1 - \frac{3 - 2f_\infty}{2(1 - f_\infty)}\, s_0\right)e^{-(3-2f_\infty)P_s N}\right)\right] \tag{7'}$$

This is a general expression, relating the expected performance of twin-contacts to that of single contacts. It is evident that the idealistic performance of $P_t = P_s^2$, i.e., the probability of a twin-contact failure is the square of that for single contacts, can only be achieved if: (a) $f_\infty = 1.0$, i.e., persistent half-opens never occur, (b) $\bar{n} = 0$, i.e., each temporary half-open occurring during one operation will clear during the subsequent operation, and (c) $s_0 = 0$, i.e., there is no initial contamination. These conditions are never obtained in practice and generally P_t is much greater than P_s^2.

Equation 7′ also indicates that at the beginning of operation, when the exponent is much less than 1.0, P_t is given by:

$$(P_t)_0 = P_s^2(1 + 2\bar{n} f_\infty) + P_s s_0 \tag{8}$$

Numerically if $\bar{n} = 50$, $f_\infty = 1.0$ and $s_0 = 0$, $P_t = 101P_s^2$ which is 101 times worse than the idealistic performance of $P_t = P_s^2$. The initial rate

of failure of twin contacts is also quite sensitive to initial contact contamination. If, for example, $s_0 = 10^{-3}$, i.e., $1/1000$ of the twin contacts are permanently half-open to start with, and for the same numbers used above and $P_s = 1/10^6$, equation 8 gives $P_t = 1.1 \times 10^{-9}$. This corresponds to an 11 fold increase in twin-contact failures just due to an initial contamination s_0 of 0.1 per cent. This performance is also 1100 times worse than the idealistic performance of $P_t = P_s^2 = 10^{-12}$.

By operation, the performance of twin contacts will exponentially deteriorate according to equation 7. It will asymptotically approach a constant rate of failure given by:

$$(P_t)_\infty = P_s^2(1 + 2\bar{n}f_\infty) + 2P_s\frac{1 - f_\infty}{3 - 2f_\infty} \tag{9}$$

This is independent of the initial contamination s_0 and is practically reached in a number of operations:

$$(n_t)_\infty = 3/(P_s(3 - 2f_\infty)) \tag{10}$$

The worst performance of twin contacts is obtained when $f_\infty = 0$, i.e.,

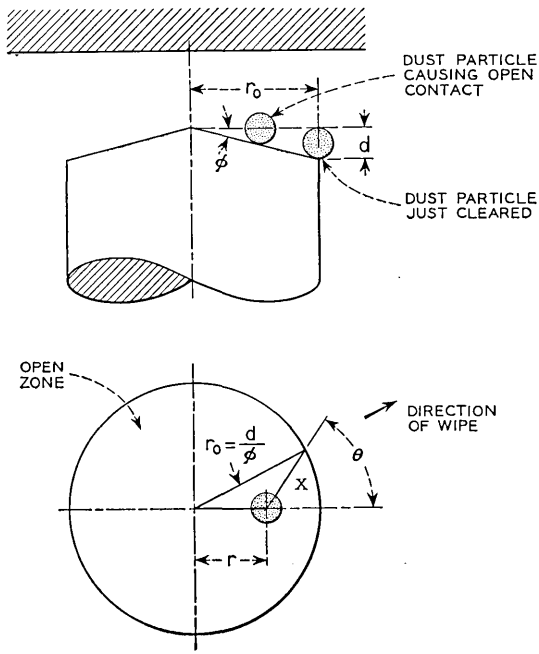

Fig. 3 — Particle in open zone.

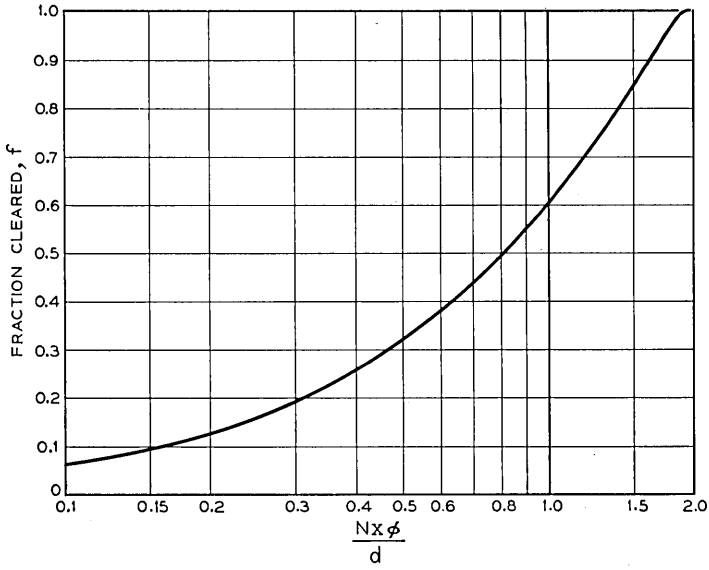

Fig. 4 — Accumulative probability distribution of fraction cleared "f" versus $(Nx\varphi/d)$.

every half-open persists indefinitely, and is given by $(P_t)_\infty = \frac{2}{3}P_s$. In other words, *by replacing single contacts by twin contacts the frequency of failures is decreased by only one third.* In practice, however, f_∞ is rarely that low and under the wide variety of central office conditions it may range between 0.85 and 0.98. This corresponds to a range of performance of $P_t = 0.23\ P_s$ to $0.038\ P_s$ or the frequency of twin-contact failures is about $\frac{1}{4}$ to $\frac{1}{26}$ that of single contacts.

CLEARING AN OPEN BY MECHANICAL OPERATION

Introduction

Fig. 3 shows a diagrammatic sketch of the contact area with spherical particles preventing metallic contact. The surface irregularity has a slope φ and the particle diameter is d. It is evident that the particle will produce an open if it falls within a limiting radius $r_0 = d/\varphi$. The area within r_0 is called the "open zone." For a particle at a radius r within the open zone, mechanical wipe w will tend to displace the particle at an angle θ, $0 \leqq \theta \leqq 2\pi$. The open is cleared when the particle is displaced a distance X sufficient to drive it out of the open zone. For unidirectional

wipe, if the average displacement per operation is x, the number of operations required to clear an open is $N = X/x$. Since, however, the initial location of the particle r has a triangular probability distribution, $0 \leqq r \leqq r_0$, and the direction of wipe θ has a rectangular distribution, $0 \leqq \theta \leqq 2\pi$, the number of operations for clearing must have a corresponding probability distribution. This has been determined graphically and is shown in Fig. 4. This is an accumulative probability distribution of the fraction cleared "f" versus $(Nx\varphi/d)$. It indicates that 50 per cent of the opens will clear at $(Nx\varphi/d) = 0.8$ and 100 per cent at 2.0. The corresponding number of operations N can be determined only if x is known* under the operating conditions. By increasing the contact wipe w one may expect x, the average displacement per operation, to increase. Also by increasing the force, the frictional driving force will increase and x should also increase. One may, therefore, tentatively assume that x is proportional to $w^a F^b$ and the distribution function may be put in the form $f = f(Nw^a F^b)$, keeping all the other parameters fixed. This suggests that by varying the contact wipe w and the contact force F, the distribu-

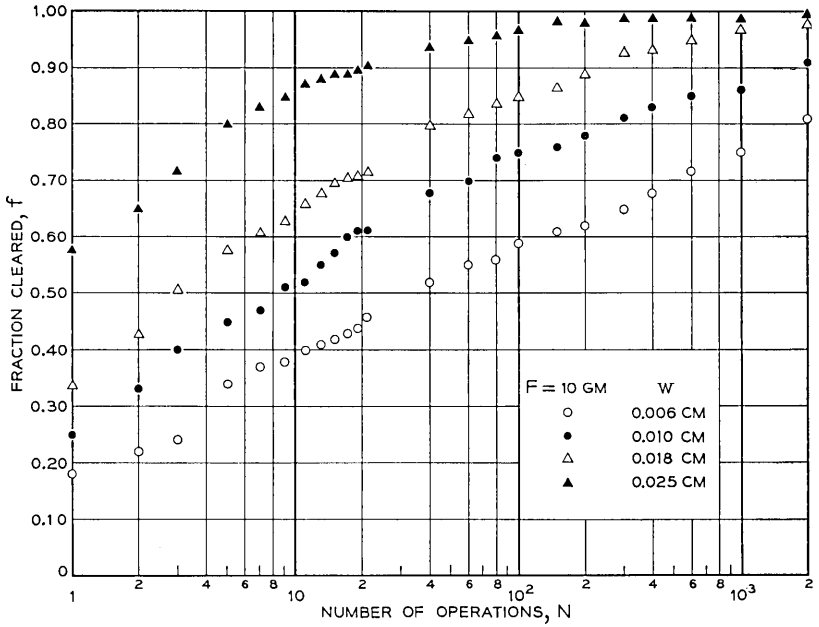

Fig. 5 — Effect of wipe.

* For a certain surface roughness φ and particle size d.

tion function f should be unique if plotted against Nw^aF^b where the a and b are constants to be determined experimentally.

Experiment

Cotton lint fibres and clean contact surfaces were used exclusively in this study. The fibres were essentially in the form of ribbons a few microns thick and of variable width and length. By using a dust separator,* lint fibres with a well controlled size distribution were collected on a glass plate. The setting used gave a rather uniform monolayer of fibres 80 per cent of which had a width between 10 and 20 microns.

The contacts tested were flat and made of palladium.† They were cleaned with methyl alcohol and distilled water, then dried. The collected lint fibres were transferred to the surface of one contact by a special adapter which allows the pressing of the contact on the glass

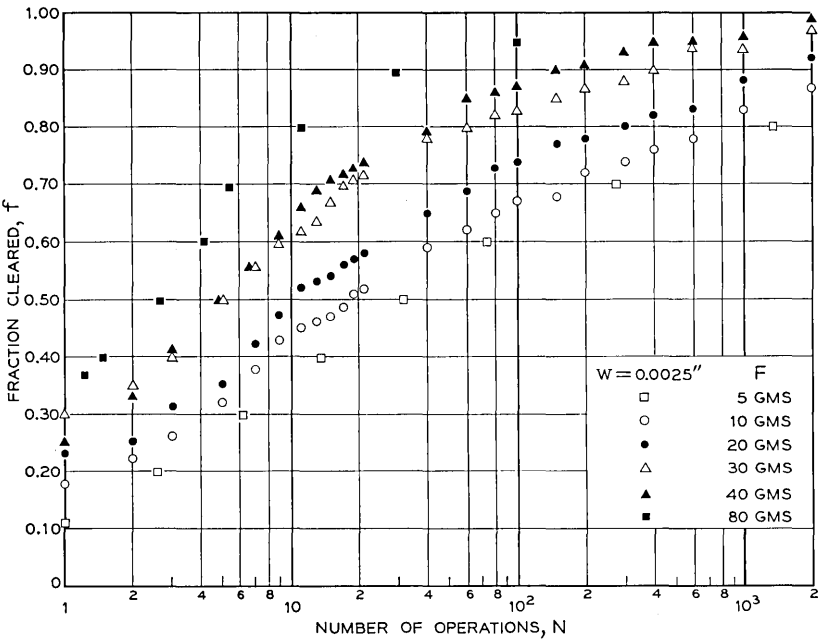

Fig. 6 — Effect of force.

* Based on controlling sedimentation by adjusting air speed in a two-stage separator.

† Contact surface roughness was controlled by frequently polishing the contact surfaces by a fixed process.

plate with the fibres. The pressing force was the same as that used in the subsequent operation of the contacts. The contacts were then operated at four operations per second in a sealed compartment. After each closure a checking circuit using 48 volts, and a maximum current of 0.50 amp., checked the continuity in the contacts. When the open was cleared the unit automatically stopped and the corresponding number of operations was obtained from a counter. The maximum number of operations allowed for each run was 2,000. For one set of operating conditions, it was necessary to repeat the above for at least 150 times before a representative clearing distribution was obtained.

Results

Effect of wipe: Fig. 5 shows the results obtained for a range of wipes between 0.006 and 0.025 cm at a constant force of 10 grams. As expected, the clearing rate was higher for larger wipes.

Effect of force: Fig. 6 was obtained at a constant wipe of 0.006 cm and a set of forces between 5 and 80 grams. Large forces gave higher clearing rates. The effect of changing the force, however, is not as significant as that of changing the wipe.

As outlined in the preceding introduction, the above data was replotted as fraction clearing f versus Nw^aF^b. The results are shown in Fig. 7 with $a = 3$ and $b = 1.0$. As indicated, the points converged to a single average line with comparatively small spread.* This shows that, at least for the range covered, the change in clearing rate obtained by changing the wipe say by a factor of two can also be obtained by changing the force by a factor of 8.

To determine the persistency $(1 - f_\infty)$, one may choose an arbitrary number of operations for defining it. If 2,000 operations is chosen, one may determine from the above data the fraction, $(1 - f_{2,000})$, that will persist to beyond 2,000 operations. This was done and the results are plotted in Fig. 8 as $(1 - f_{2,000})$ versus $F^{1/3}w$. This suggests the following relation:

$$(1 - f_{2,000}) = e^{-F^{1/3}w/0.006} \tag{11}$$

where F is in grams and w in cms. This expression allows the determination of the effects of force and wipe on the performance of twin contacts by substituting in Equations 7′ through 10. Similarly the average number of operations $\bar{n}$, used in the above equations, may be obtained. This may

* This same convergence was obtained, but not presented here, by plotting f versus NF at constant w and f versus Nw^3 at constant F.

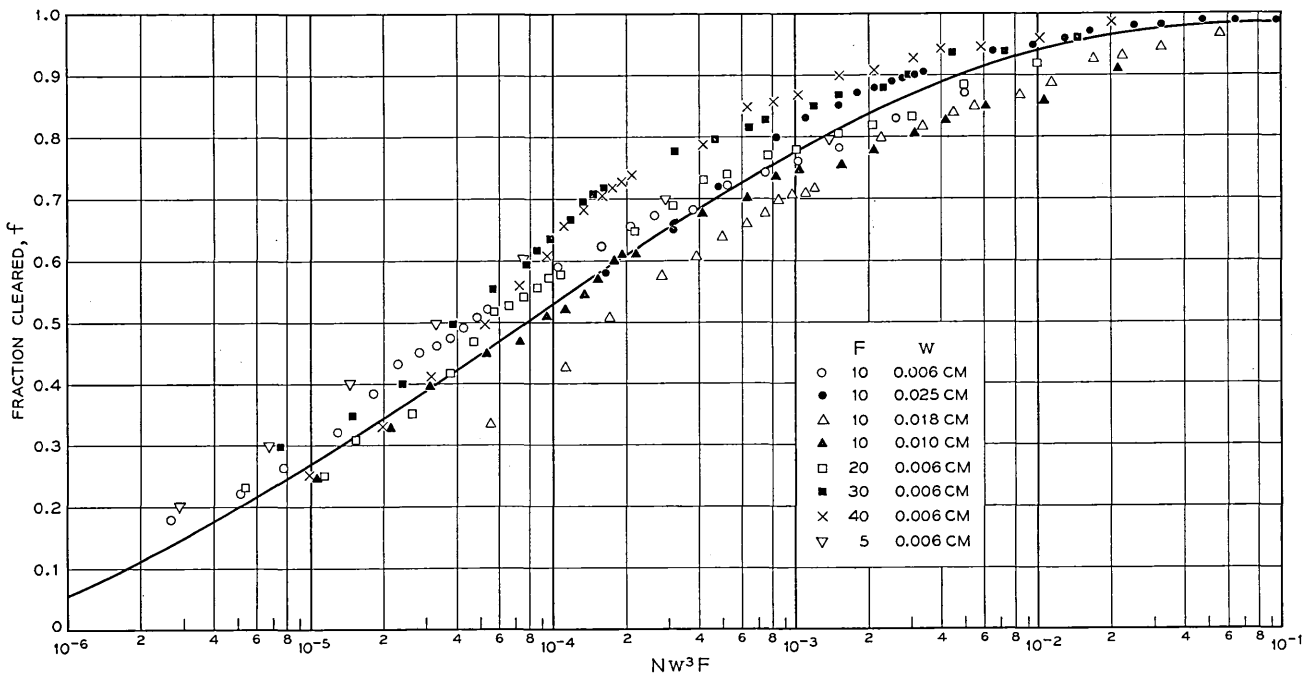

Fig. 7 — Average experimental curve; fraction cleared "f" versus Nw^3F.

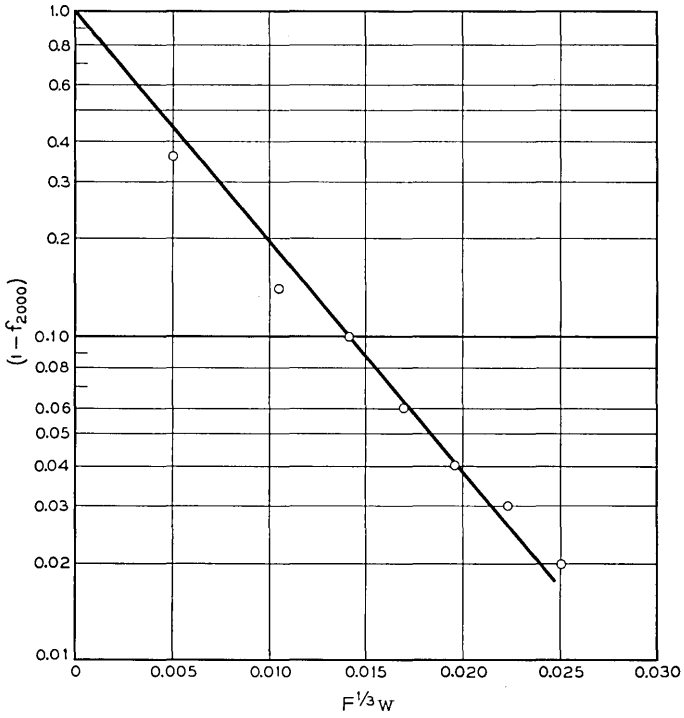

Fig. 8 — Persistency of opens at 2,000 operations.

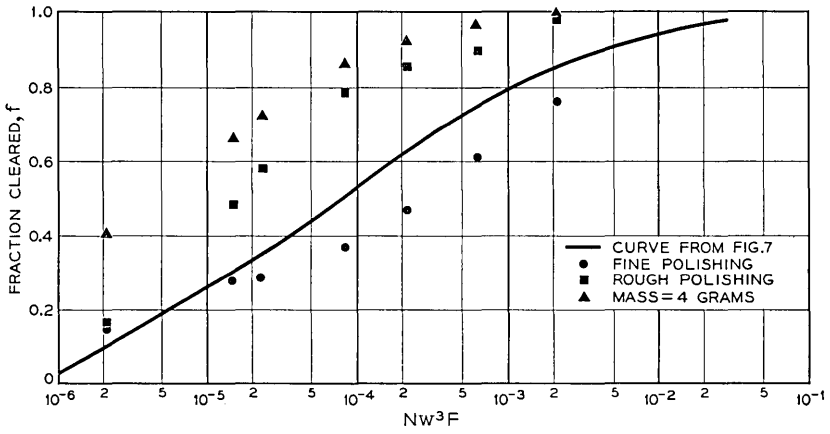

Fig. 9 — Effect of other parameters.

be arbitrarily defined as the number of operations at which 50 per cent of the opens will have cleared. At this or any other value of f one obtains from Fig. 7 $\bar{n}Fw^3$ = constant or $\bar{n}$ is inversely proportional to Fw^3.

OTHER EFFECTS

Fig. 9 shows the results obtained by varying other parameters. The solid line, obtained from Fig. 7 is shown for comparison. Indicated are the effects of fine polishing and rough polishing of the contact surfaces and of increasing the mass of the moving contact from 0.5 to 4 grams.

Bell System Technical Papers Not Published in this Journal

AHEARN, A. J., see HANNAY, N. B.

BEACH, A. L., see GULDNER, W. G.

BIDDULPH, R.[1]

Short Term Autocorrelation Analysis and Correlatogram of Spoken Digits, J. Acous. Soc. Am., **26,** pp. 539–541, July, 1954.

BRATTAIN, W. H., see GARRETT, C. G. B.

BULLINGTON, K.[1]

Reflection Coefficients of Irregular Terrain, Proc. I.R.E., **42,** pp. 1258–1262, Aug., 1954.

DEWALD, J. F.,[1] and LEPOUTRE, GERARD[2]

The Thermoelectric Properties of Metal — Ammonia Sodium and Potassium at −33°, J. Am. Chem. Soc., **76,** pp. 3369–3373, July, 1954.

FINE, M. E.,[1] and KENNEY, NANCY T.[1]

Moduli and Internal Friction of Magnetite as Affected by the Low-Temperature Transformation, Phys. Rev., **94,** pp. 1573–1576, June 15, 1954.

---

[1] Bell Telephone Laboratories, Inc.
[2] Faculté Libre des Sciences, Lille, Nord, France.

FLETCHER, R. C.,[1] YAGER, W. A.,[1] PEARSON, G. L.,[1] and MERRITT, F. R.[1]

Hyperfine Splitting in Spin Resonance of Group V Donors in Silicon, Letter to the Editor, Phys. Rev., **95,** pp. 844–5, Aug. 1, 1954.

GAMBRELL, J. B., JR.[1]

What is a Printed Publication Within the Meaning of the Patent Act, J. Patent Office Society, **36,** pp. 391–405, June, 1954.

GARRETT, C. G. B.,[1] and BRATTAIN, W. H.[1]

Self Powered Semiconductor Amplifier, Letter to the Editor, Phys. Rev., **95,** pp. 1091–1092, Aug. 15, 1954.

GREEN, E. I.[1]

Creative Thinking in Scientific Work, Elec. Eng., **73,** pp. 489–494, June, 1954.

GREMLING, R. C., see WRIGHT, MARIE G.

GULDNER, W. G.,[1] and BEACH, A. L.[1]

Gasometric Method for Determination of Hydrogen in Carbon, Analytical Chemistry, **26,** pp. 1199–1202, July, 1954.

HANNAY, N. B.[1]

A Mass Spectrograph for the Analysis of Solids, Rev. Scient. Instr., **25,** pp. 644–648, July, 1954.

HANNAY, N. B.,[1] and AHEARN, A. J.[1]

Mass Spectrographic Analysis of Solids, Analytical Chemistry, **26,** pp. 1056–1058, June, 1954.

HERRING, CONYERS[1]

Pole of Low Energy Phonons in Thermal Conduction, Phys. Rev., **95,** pp. 954–965, Aug. 15, 1954.

[1] Bell Telephone Laboratories, Inc.

HOGAN, C. L., see VAN UITERT, L. G.

KARLIN, J. E., see MUNSON, W. A.

KELLY, H. P.[1]

Differential Phase and Gain Measurements in Color Television Systems, Elec. Eng., **73,** pp. 799–802, Sept., 1954.

KELLY, H. P.[1]

Color Video Tester Checks Distortion, Electronics, **27,** pp. 128–131, Sept., 1954.

KENNEY, NANCY T., see FINE, M. E.

KING, R. A.,[1] and MORGAN, S. P.[1]

Transmission Formulas and Charts for Laminated Coaxial Cables, Proc. I.R.E., **42,** pp. 1250–1258, Aug., 1954.

KISLIUK, PAUL[1]

Arcing at Electrical Contacts on Closure — Part V. The Cathode Mechanisms of Extremely Short Arcs, J. Appl. Phys., **25,** pp. 897–900, July, 1954.

LEGG, V. E.,[1] and OWENS, C. D.[1]

Magnetic Ferrites: New Materials for Modern Applications, Elec. Eng., **73,** pp. 726–729, August, 1954.

LEPOUTRE, GERARD, see DEWALD, J. F.

LOVELL, L. C., see VOGEL, F. L.

MAITA, J. P., see MORIN, F. J.

MASON, D. R.[1]

Considerations on Chemical Engineering Design Problems (in French); Chimie et Industrie, **71,** pp. 477–481, March, 1954.

[1] Bell Telephone Laboratories, Inc.

McHugh, K.[3]

Speculation on the Failure of Telephony, Telephony, 147, pp. 17–19, 42–43, Aug. 14, 1954.

Merritt, F. R., see Fletcher, R. C.

Merz, W. J.[1]

Domain Formation and Domain Wall Motions in Ferroelectric $BaTiO_3$ Single Crystals, Phys. Rev., **95,** pp. 690–698, Aug. 1, 1954.

Morgan, S. P., see King, R. A.

Morin, F. J.,[1] and Maita, J. P.[1]

Conductivity and Hall Effect in the Intrinsic Range of Germanium, Phys. Rev., **94,** pp. 1525–1529, June 15, 1954.

Munson, W. A.,[1] and Karlin, J. E.[1]

The Measurement of Human Channel Transmission Characteristics, J. Acous. Soc. Am., **26,** pp. 542–553, July, 1954.

Owens, C. D., see Legg, V. E.

Pearson, G. L., see Fletcher, R. C.

Read, W. T., see Vogel, F. L.

Schafer, J. P., see Van Uitert, L. G.

Schawlow, A. L.[1]

Nuclear Quadrupole Resonances in Solid Bromine in Iodine Compounds, Chem. Phys., **22,** pp. 1211–1214, July, 1954.

Shockley, W., see van Roosbroeck, W.

---

[1] Bell Telephone Laboratories, Inc.
[3] New York Telephone Company.

TIEN, PING KING[1]

Bifilar Helix for Backward Wave Oscillators, Proc. I.R.E., **42,** pp. 1137–1143, July, 1954.

VAN ROOSBROECK, W.,[1] and SHOCKLEY, W.[1]

Photo-Radiative Recombination of Electrons and Holes in Germanium, Phys. Rev., **94,** pp. 1558–1560, June 15, 1954.

VAN UITERT, L. G.,[1] SCHAFER, J. P.,[1] and HOGAN, C. L.[1]

Low Loss Ferrites for Applications at 4,000 Millicycles per Second, Letter to the Editor, J. Appl. Phys., **25,** p. 925, July, 1954.

VOGEL, F. L.,[1] READ, W. T.,[1] and LOVELL, L. C.[1]

Recombination of Holes and Electrons at Lineage Boundaries in Germanium, Letter to the Editor, Phys. Rev., **94,** pp. 1791–1792, June 15, 1954.

WALKER, A. C.[1]

Hydrothermal Growth of Quartz Crystals, Ind. and Eng. Chem., **48,** pp. 1670–1676, Aug., 1954.

WOLFF, P. A.[1]

Theory of Secondary Electron Cascade in Metals, Phys. Rev., **95,** pp. 56–66, July 1, 1954.

WRIGHT, MARIE G.,[1] and GREMLING, R. C.[1]

Xerographic Short Cut, Special Libraries, **45,** pp. 250–251, July–August, 1954.

YAGER, W. A., see FLETCHER, R. C.

[1] Bell Telephone Laboratories, Inc.

Recent Monographs of Bell System Technical Papers Not Published in This Journal*

ANDERSON, P. W., MERRITT, F. R., REMEIKA, J. P., and YAGER, W. A.

Magnetic Resonance in α Fe_2O_3, Monograph 2229.

BARSTOW, J. M., and CHRISTOPHER, H. N.

Measurement of Random Monochrome Video Interference, Monograph 2259.

BOND, W. L.

Notes on Solution of Problems in Odd Job Vapor Coating, Monograph 2302.

CHRISTOPHER, H. N., see BARSTOW, J. M.

CLARK, M. A.

An Acoustic Lens as a Directional Microphone, Monograph 2291.

CLOGSTON, A. M., and HEFFNER, H.

Focusing of an Electron Beam by Periodic Fields, Monograph 2267.

CORY, S. I.

A New Portable Telegraph Transmission Measuring Set, Monograph 2248.

DARROW, K. K.

Solid State Electronics, Monograph 2253.

* 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.

DITZENBERGER, J. A., see FULLER, C. S.

FINE, M. E., and KENNEY, NANCY T.

Moduli and Internal Friction of Magnetite as Affected by the Low-Temperature Transformation, Monograph 2251.

FRACASSI, R. D., and KAHL, H.

Type-ON Carrier Telephone, Monograph 2296.

FULLER, C. S., STRUTHERS, J. D., DITZENBERGER, J. A., and WOLFSTIRN, K. B.

Diffusivity and Solubility of Copper in Germanium, Monograph 2270.

GALT, J. K., YAGER, W. A., and MERRITT, F. R.

Temperature Dependence of Ferromagnetic Resonance Line Width in a Nickel Iron Ferrite: A New Loss Mechanism, Monograph 2245.

GILBERT, E. N.

Lattice Theoretic Properties of Frontal Switching Functions, Monograph 2246.

GREEN, E. I.

The Decilog: A Unit for Logarithmic Measurement, Monograph 2254.

HAGSTRUM, H. D.

Instrumentation and Experimental Procedure for Studies of Electron Ejection by Ions and Ionization by Electron Impact, Monograph 2256.

HEFFNER, H.

Analysis of the Backward-Wave Traveling-Wave Tube, Monograph 2285.

HEFFNER, H., see CLOGSTON, A. M.

JAFFE, H., see MASON, W. P.

JOHNSON, J. B., and MCKAY, K. G.

Secondary Electron Emission From Germanium, Monograph 2279.

KAHL, H., see FRACASSI, R. D.

KENNEY, NANCY, T., see FINE, M. E.

KOMPFNER, R., and WILLIAMS, N. T.

Backward-Wave Tubes, Monograph 2295.

KRETZMER, E. R.

An Amplitude Stabilized Transistor Oscillator, Monograph 2239.

LEWIS, H. W.

Search for the Hall Effect in a Superconductor I. — Experiment, Monograph 2255.

LINVILL, J. G.

RC Active Filters, Monograph 2260.

LINVILL, J. G.

Transistor Negative-Impedance Converters, Monograph 2294.

MAITA, J. P., see MORIN, F. J.

MASON, W. P., and JAFFE, H.

Methods for Measuring Piezoelectric, Elastic, and Dielectric Coefficients of Crystals and Ceramics, Monograph 2241.

MCKAY, K. G., see JOHNSON, J. B.

MENDEL, J. T., QUATE, C. F., and YOCUM, W. H.

Electron Beam Focusing With Periodic Permanent Magnet Fields, Monograph 2240

MERRITT, F. R., see ANDERSON, P. W.

MERRITT, F. R., see GALT, J. K.

MORIN, F. J., and MAITA, J. P.

Conductivity and Hall Effect in the Intrinsic Range of Germanium, Monograph 2300.

MORIN, F. J., see PEARSON, G. L.

PEARSON, G. L., READ, W. T., JR., and MORIN, F. J.

Dislocations in Plastically Deformed Germanium, Monograph 2230.

PFANN, W. G.

Redistribution of Solutes by Formation and Solidification of a Molten Zone, Monograph 2290.

PIERCE, J. R.

Coupling of Modes of Propagation, Monograph 2252.

PRINCE, M. B.

Drift Mobilities in Semiconductors I, Germanium II, and Silicon, Monograph 2271.

QUATE, C. F., see MANDEL, J. T.

READ, W. T., JR., see PEARSON, G. L.

REMEIKA, J. P.

Method for Growing Barium Titanate Single Crystals, Monograph 2247.

REMEIKA, J. P., see ANDERSON, P. W.

RYDER, R. M., and SITTNER, W. R.

Transistor Reliability Studies, Monograph 2263.

SHIVE, J. N., see SLOCUM, A.

SHOCKLEY, W., see VAN ROOSBROECK, W.

SITTNER, W. R., see RYDER, R. M.

SLOCUM, A., and SHIVE, J. N.

Shot Dependence of P-N Junction Phototransistor Noise, Monograph 2265.

SMITH, C. S.

Piezoresistance Effect in Germanium and Silicon, Monograph 2233.

SNOKE, L. R.

Observations on a Possible Method of Predicting Soil-Block Bioassay Thresholds by Distillation Characteristics of the Weathered Creosotes, Monograph 2238.

THOMAS, D. E.

A Point-Contact Transistor VHF FM Transmitter, Monograph 2234.

VALDES, L. B.

Resistivity Measurements on Germanium for Transistors, Monograph 2261.

VAN ROOSBROECK, W., and SHOCKLEY, W.

Photon-Radiative Recombination of Electrons and Holes in Germanium, Monograph 2306.

VARNEY, R. N.

Liberation of Electrons by Positive-Ion Impact on the Cathode of a Pulsed Townsend Discharge Tube, Monograph 2232.

WALKER, L. R.

Stored Energy and Power Flow in Electron Beams, Monograph 2264.

WICK, R. F.

Solution of the Field Problem of the Germanium Gyrator, Monograph 2301.

WILLIAMS, N. T., see KOMPFNER, R.

WOLFF, P. A.

Theory of Plasma Waves in Metals, Monograph 2262.

WOLFSTRIN, K. B., see FULLER, C. S.

YAGER, W. A., see ANDERSON, P. W., and GALT, J. K.

YOCOM, W. H., see MENDEL, J. T.

Contributors to this Issue

M. M. Atalla, B.S., Cairo University, 1945; M.S., Purdue University, 1947; Ph.D., Purdue University, 1949; Studies at Purdue undertaken as the result of a scholarship from Cairo University for four years of graduate work. Bell Telephone Laboratories, 1950–. For the past three years he has been a member of the Switching Apparatus Development Department, in which he is supervising a group doing fundamental research work on contact physics and engineering. Current projects include fundamental studies of gas discharge phenomena between contacts, their mechanisms, and their physical effects on contact behavior; also fundamental studies of contact opens and resistance. In 1950, an article by him was awarded first prize in the junior member category of the A.S.M.E. He is a member of Sigma Xi, Sigma Pi Sigma, Pi Tau Sigma, the American Physical Society, and an associate member of the A.S.M.E.

Rosemary E. Cox, B.A., Ladycliff College, 1949; M.S. Fordham University, 1950; Bell Telephone Laboratories, 1951–. Miss Cox, who taught high school mathematics for a year before coming to the Laboratories, has been engaged in fundamental studies of contact physics. She won a New York State University Scholarship and scholarships from Ladycliff and Fordham.

Gerald V. King, B.S., Carnegie Institute of Technology, 1920; Western Electric Company, 1921–24; Bell Telephone Laboratories, 1925–. Mr. King analyzed customer orders for step-by-step and manual systems for three years before turning to the design and checking of manual and dial PBX and community dial offices. From 1932 to 1939 he engaged in fundamental studies and development of new local and toll crossbar systems. He was involved in military work from 1939 to 1944 and since then he has been concerned with design and development of AMA accounting centers, central offices and crossbar tandem systems. He was appointed Switching Systems Development Engineer in 1952.

Stewart E. Miller, University of Wisconsin, 1936–39; B.S. and M.S., Massachusetts Institute of Technology, 1941. Bell Telephone

Laboratories, 1941–. Since June 1954, Mr. Miller has been Assistant Director of Radio Research at Holmdel and has been in charge of research on guided systems and associated millimeter and microwave techniques. During World War II, he worked on airborne radar systems. He also worked on coaxial carrier transmissions systems. Mr. Miller holds patents in connection with automatic frequency control, an oscillator control scheme and the D-C amplifier. Member of the I.R.E., Eta Kappa Nu, Tau Beta Pi and Sigma Xi.

NORMAN A. NEWELL, E. E., Lehigh University, 1920. Mr. Newell joined Bell Telephone Laboratories in 1934 after fourteen years with the American Telephone and Telegraph Company. He has formed requirements for the development of circuit and equipment arrangements for intertoll trunk signaling systems, local and toll manual switchboards, toll switchboard trunking systems and automatic toll systems. He helped coordinate these projects and prepared descriptions of the systems for the operating companies. Mr. Newell holds patents in the fields of straightforward trunking, toll-call timing and single-frequency signaling. He is a member of the A.I.E.E., Tau Beta Pi and Pi Delta Epsilon.

WILLIAM PFERD, B.S. in M.E. Rutgers University, 1947; M.S. in M.E., Newark College of Engineering 1951. Mr. Pferd has recently been concerned with coin collector development. Previously, he worked on design and development of the station ringer for the 500-type telephone set and the dial mechanism for the same set. During World War II he served as a photographic intelligence officer in Italy with the 98th Bomb Group.

JOHN R. PIERCE, B.S., M.S., Ph.D., California Institute of Technology, 1933, 1934 and 1936; Bell Telephone Laboratories, 1936–; Appointed Director of Electronics Research, 1952. Dr. Pierce has specialized in the development of electron tubes and microwave research since joining the Laboratories. During the war he concentrated on the development of electronic devices for the armed forces. Since the war he has done research leading to the development of the beam traveling-wave tube for which he was awarded the 1947 Morris Liebmann Memorial Prize of the Institute of Radio Engineers. Dr. Pierce is the author of two books: *Theory and Design of Electron Beams*, published in second edition this year, and *Traveling Wave Tubes* (1950). He was voted the "Outstanding Young Electrical Engineer of 1942" by Eta Kappa Nu. Member of the A.I.E.E., Tau Beta Pi and Eta Kappa Nu. Fellow of the American Physical Society and the I.R.E.

ALLAN WEAVER, B.S. in E. E., University of Nebraska, 1921. Since 1945 Mr. Weaver has been in charge of a group concentrated on toll signaling with particular attention to the development of single-frequency signaling for use in connection with nationwide dialing systems. He joined Bell Telephone Laboratories in 1934 after thirteen years with American Telephone and Telegraph Company and was first concerned with the development of telegraph, telephotograph and teletypewriter systems. During World War II he was assigned to radar development. He holds thirty-seven patents in the fields of telegraphy, teletypography and signaling. Mr. Weaver is a member of the A.I.E.E., I.R.E. and Sigma Xi.