The Bell System Technical Journal

Vol. XV *April, 1936* *No. 2*

The Reliability of Short-Wave Radio Telephone Circuits *

By R. K. POTTER and A. C. PETERSON, JR.

From empirical measurements of noise-to-signal ratio made during the routine operation of short-wave radio telephone circuits there is obtained a general relation between percentage lost circuit time and transmission improvement in decibels. In this relation "percentage lost circuit time" is the percentage of time that the noise-to-signal ratio is considered unsatisfactory. No attempt is made to define such a standard quantitatively.

If, from past experience with a long-range, short-wave telephone, telegraph or broadcast circuit, it is known that the circuit is unsatisfactory a certain percentage of the time, the above-mentioned relation may be used to estimate the effect of transmission improvement upon this percentage of unsatisfactory or lost time. For a given circuit the variation in percentage lost circuit time, as the standard for the tolerable service is changed by a given number of decibels, may also be estimated.

There are included estimates of the relation between the number of lost time intervals of various lengths and transmission improvement.

INTRODUCTION

WITHIN a comparatively few years short-wave radio telephone circuits have become an important part of the international communication network. These years have represented a wide variety of experience ranging between the quiet and the disturbed extremes of an eleven-year sunspot cycle. An attempt is made here to review some of this transmission experience in a quantitative way and show certain relations that may be useful in the engineering of short-wave circuits.

During a magnetically disturbed year, such as 1930, a low-power short-wave transmitter with a simple antenna arrangement would have provided very uncertain means for communication across the North Atlantic. The percentage of time that such equipment could transmit what according to lenient standards in terms of noise-to-signal ratio are useful telephone signals, would have been very low. If the power of the transmitter were increased or a directive antenna employed to reduce the noise-to-signal ratio the percentage useful time would, as based upon the same standards, be increased or conversely the percentage lost time decreased. Any improvement which will decrease

*To be presented at joint meeting of U.R.S.I. and I.R.E., Washington, D. C., May 1, 1936.

the noise-to-signal ratio at the receiver output will accomplish a certain increase in usefulness of the circuit.

It is difficult to set down noise-to-signal ratios that may be employed to distinguish between satisfactory and unsatisfactory service. Such requirements would be different in the cases of telephone, telegraph and broadcast circuits. They would also depend to a considerable extent upon the facilities that it is technically and economically reasonable to provide. During times of magnetic disturbance radio telephone circuits are continued in service when noise conditions are very appreciably worse than would be tolerated on wire telephone circuits. In this emergency situation it is, of course, necessary to maintain service on the radio links as long as communication can be carried on with a reasonable degree of satisfaction.

Without a quantitative definition of the boundary between satisfactory and unsatisfactory service in terms of noise-to-signal ratio it is possible to determine from an analysis of past operating experience what percentage of the time a certain circuit was unsatisfactory. With such information available it would be useful to know how much this percentage could be reduced by the application of transmission improvements. There is developed below a form of "reliability" curve that makes it possible to estimate approximately the effect of such transmission improvements in terms of decibels upon the percentage of unsatisfactory or lost circuit time.

As a background for the following discussion it will be helpful to review briefly the conditions experienced on a typical short-wave circuit and the way in which these are related to the present analysis. For example, the instability of short-wave transmission over the North Atlantic path is well known. There are days when these trans-atlantic short-wave signals are remarkably good and others during times of magnetic disturbance when they are exceptionally poor. Between these two extemes is a wide range of circuit conditions. The situation is illustrated in idealized fashion by Fig. 1. The ordinates here represent average noise-to-signal ratios as measured on successive days at the receiver output and curve A of Fig. 1 (a) shows how this average might vary over an interval of many days. A certain noise-to-signal ratio such as is indicated by the horizontal line B might be specified as the highest value tolerable for a useful circuit according to some predetermined standard. Then the width of the cross-hatched intervals C represents the lost circuit time.

Fig. 1 (b) is the same as 1 (a) except that here a transmission improvement of x db has been applied so that the noise-to-signal ratio at the receiver output is on all days reduced x db and the curve A is

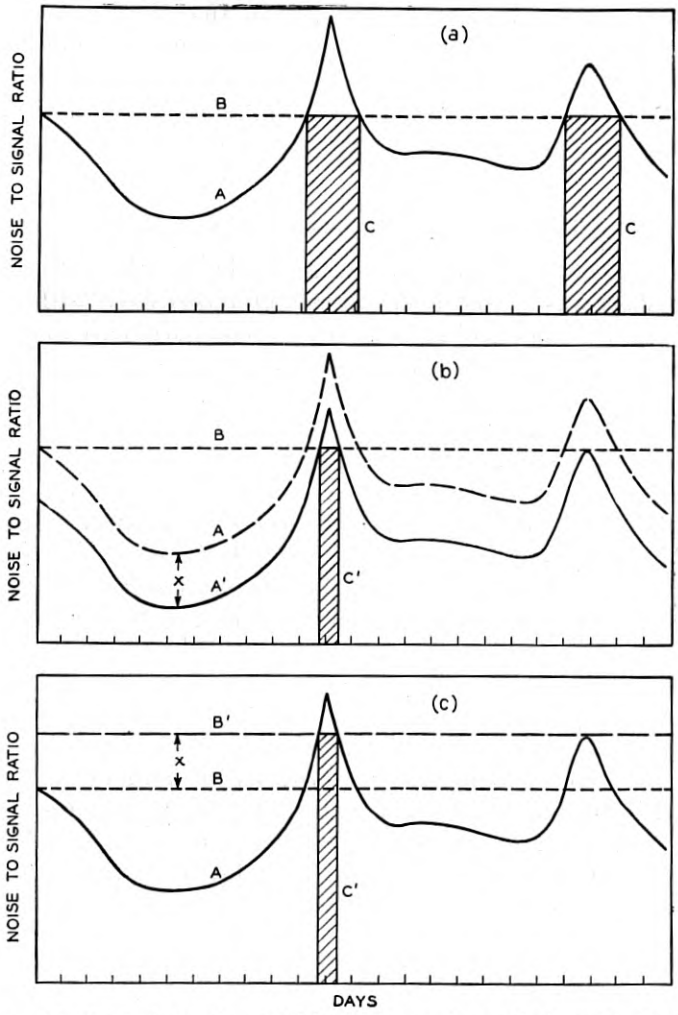


FIG. 1—Idealized illustrations of (*a*) the day-to-day variation of noise-to-signal ratio and its effect upon lost circuit time, (*b*) the effect upon lost circuit time of transmission improvements and (*c*) the effect upon lost circuit time of a change in the maximum tolerable noise-to-signal ratio.

consequently lowered to position A'. Assuming the noise-to-signal requirement B remains the same the lost time is reduced to the interval C'. If instead of applying a transmission improvement we increase the tolerable noise-to-signal requirement for a useful circuit or degrade the standard requirement by x db in the case of Fig. 1 (a), the horizontal line B is shifted upward x db as shown in Fig. 1 (c) and the lost time corresponds to that for the case of x db transmission improvement in Fig. 1 (b).

From the above illustration it is evident that if we know how the noise-to-signal ratio varies on a given circuit with a certain terminal arrangement, it is possible to determine the percentage lost circuit time for an improved or degraded system, the relative effectiveness of which can be expressed in terms of db above or below the initial arrangement.

Transatlantic Noise-to-Signal Data

As a part of the regular operating routine on the various trans-oceanic short-wave radio telephone circuits associated with the Bell System, measurements of noise at the receiver output are made at approximately half-hourly intervals. These measurements are made at a point in the voice-frequency wire circuits where the speech volume is normally held constant. They are therefore effectively measurements of noise-to-signal ratio although not expressed in such terms. The instrument used is known as the Western Electric 6-A Transmission Measuring set.[1] In the following discussion measurements made with this instrument are referred to as "6-A Noise."

In Fig. 2 (a) the upper curve shows the percentage distribution of 6-A noise values measured at New York during 1930 on an 18-mc. London-New York circuit. These and the curves to follow are plotted to an arithmetical probability scale. The year 1930 was severely disturbed and from the radio transmission standpoint is perhaps representative of the peak of the well-known eleven-year magnetic disturbance cycle. Since the performance of a two-way telephone circuit depends upon transmission conditions in the two directions the upper curve of Fig. 2 (a) does not accurately portray the full effect of the noise factor upon the circuit. The lower curve in this figure represents the distribution of the higher of simultaneous[2] 6-A noise values measured at New York and London. The small difference between the two distributions is evidence that the most important influence—that of magnetic disturbance—affects the transmission in both directions coincidentally.

It will be noted that these 6-A noise curves of Fig. 2 (a) and of the following figures bend downward in the region of low 6-A noise and upward where the 6-A noise becomes high. There is reason to believe that these bends are introduced by the terminal equipment and that the actual noise distribution of interest here approaches a straight line on the probability scale used, or in other words is a fortuituous

[1] L. Espenschied, "Methods for Measuring Interfering Noises," *Proc. I.R.E.*, Vol. 19, p. 1951, November, 1931.

[2] Measurements less than seven minutes apart at the two ends of the circuit were treated as "simultaneous" in this analysis.

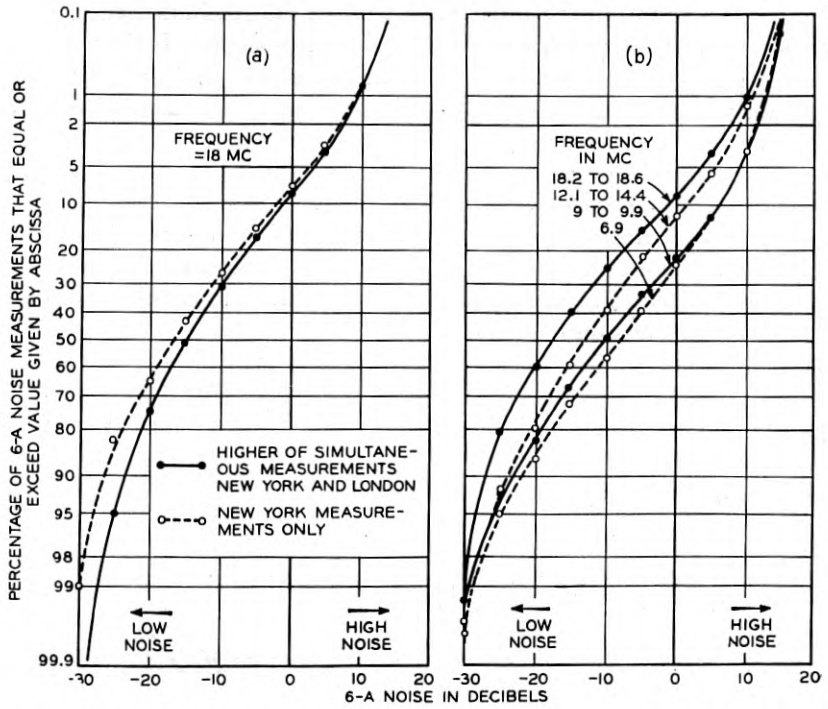


FIG. 2—Percentage distribution of 6-A noise measurements made on London to New York short-wave circuits during 1930 on (*a*) one 18-megacycle circuit and in (*b*) each of four frequency ranges.

type of distribution. The bend at the low noise end is apparently due to noise transmitted from the distant terminal and that introduced by the local receiver. These sources of noise are approximately constant and, when the atmospheric noise is very low, become the limiting factors. The bend at the high noise end of the curve is probably due to the action of the automatic volume control in the receiver. This control normally holds the speech volume approximately constant, but when noise is exceedingly high the noise in itself reduces the receiver gain and depresses both the noise and signal output. Since at such times speech volume cannot be accurately checked, the measurement is no longer an accurate indication of noise-to-signal ratio and the curve reaches a limiting value. Evidence confirming the inaccuracy of the 6-A noise readings at the high and low noise extremes will be discussed later in connection with the observed distribution of high-frequency signal intensity values.

In Fig. 2 (b) are shown distribution curves for 6-A noise values measured at New York during 1930 on several of the London to New

York circuits within the different frequency ranges indicated. All of these curves have roughly the same mid-range slope. Incidentally, if correction is made for relative transmitter power and antenna gains the horizontal separation is a direct comparison of transmission effectiveness on the different frequencies within their period of use. When such a correction is applied to the curves of Fig. 2 (b) the mid-range separation becomes less than 3 db, indicating that with equal transmitter power, antenna gains and other terminal improvements the average 6-A noise distribution is substantially the same for all times of the day when suitable frequencies are employed to cover the diurna range of transmission requirements.

Comparison of 1930, 1932 and 1934

As previously mentioned, short-wave transmission conditions were severely disturbed during the year 1930. By 1932, conditions had become much more favorable and 1934 was perhaps typical of a quiet year. In Fig. 3 (a) are included 6-A noise distribution curves for the

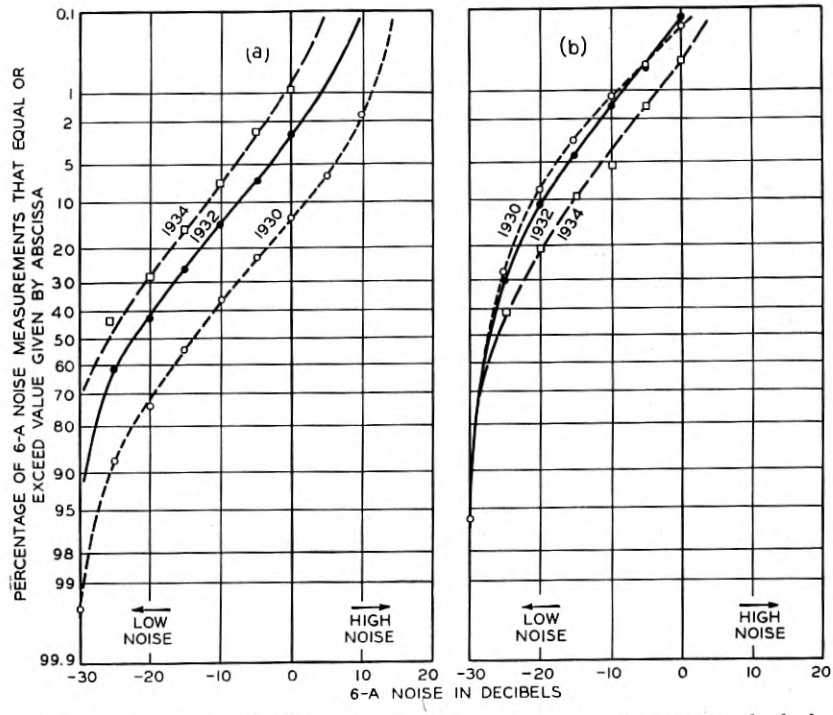


Fig. 3—Percentage distribution curves of 6-A noise measurements made during 1930, 1932 and 1934 on (*a*) all London to New York short-wave circuits and on (*b*) all Buenos Aires to New York circuits.

London-New York short-wave telephone circuits during 1930, 1932 and 1934. These three curves have very nearly the same mid-range slope, showing that the general character of the distribution was the same over the wide range of transmission conditions experienced within this interval. To those familiar with transatlantic short-wave transmission during 1930, the year 1934 would rate as comparatively undisturbed, and yet these curves indicate that for an equal percentage of measurements or, as will be apparent later, for equal lost circuit time during these two years the difference in required transmission effectiveness would be only 13 or 14 db.

Fig. 3 (b) shows that there was relatively small db separation between the 6-A noise distributions for 1930, 1932 and 1934 on the low latitude South American circuits but that the position of the curves is reversed, 1930 being better than 1934. An examination of field intensity data for these years indicates that this is due to a change in noise rather than to a change in signal transmission.

Fig. 4 (a) compares the distributions for the circuits Buenos Aires–

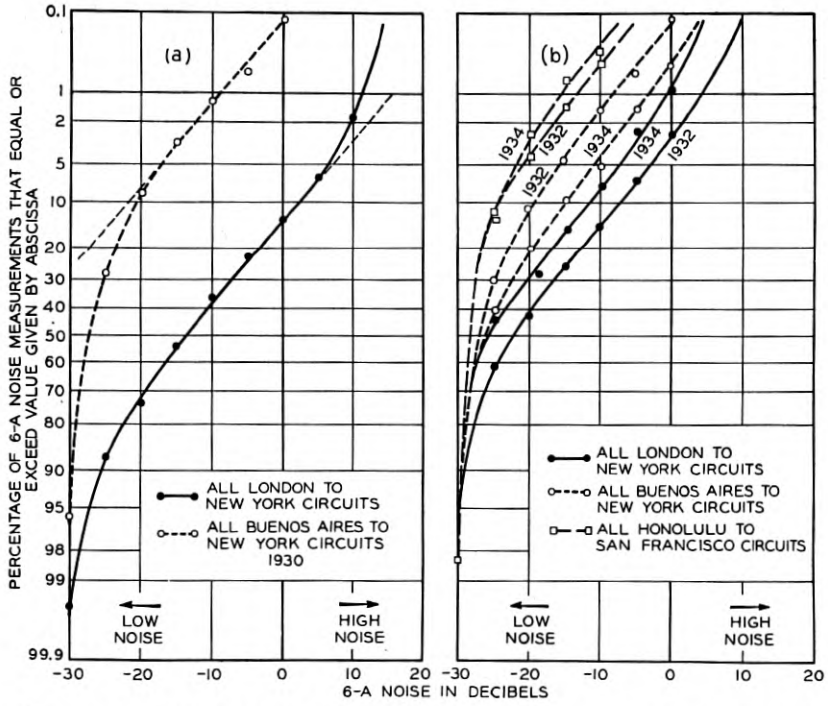


FIG. 4—Percentage distribution curves of 6-A noise measurements made on (*a*) all London to New York, and Buenos Aires to New York short-wave circuits during 1930 and on (*b*) all London to New York, Buenos Aires to New York, and Honolulu to San Francisco short-wave circuits during 1932 and 1934.

New York and London–New York for the year 1930. Again the mid-range slope as shown by the extended broken lines is about the same. The horizontal separation of roughly 25 db illustrates the much more favorable noise-to-signal conditions on the low latitude circuit, since the transmitter power and antenna gains were substantially the same in the two cases.

In Fig. 4 (b) are shown 6-A noise distribution curves for the Honolulu–San Francisco, Buenos Aires–New York and London–New York circuits representing conditions during 1932 and 1934. The mid-range slope is remarkably similar for all six curves. A comparison of these curves illustrates the high degree of reliability obtained on the circuit to Honolulu.

In Figs. 4 (a) and (b) the data for the low latitude paths cover about 9 hours of daylight operation as compared with 24-hour full time operation for the transatlantic case. Although data are not available for a 24-hour comparison some available experience indicates that such a comparison would not have altered the separation between the curves shown very appreciably. After all, the interest here is mainly in the slopes of these curves, and there is no reason to believe that the slopes would be affected.

Field Intensity Distribution

So far the discussion has shown that over the dependable portion of the 6-A noise distribution curves representative of both different circuits and different years the slopes appear to be nearly the same. If this is the case one curve may be constructed to represent approximately the effect of transmission improvement or degradation upon the performance of any long range short-wave circuit. The useful range of this curve is, however, limited by the dependable range of the 6-A noise measurements. To extend the useful range of the curve in order to estimate the effect of large changes in transmission improvement it is necessary to resort to a correction for the bend at the high noise end. Correction at the low noise end would concern the less important case of transmission degradation. Although it is possible to apply an approximate correction for the above-mentioned effect of the automatic gain control upon the bend at the high noise end it is probably more accurate to consider the distribution of field intensity data at these times of high noise-to-signal ratio.

The limiting conditions on short waves are predominantly those accompanying magnetic disturbances when the signal fields drop to very low values. The indications are that the atmospheric noise fields also decrease to a less noticeable extent during these disturb-

ances,[3] so that if this were the only effect to be considered the 6-A noise which is dependent upon the noise-to-signal ratio would increase slowly. But first circuit and tube noise in the receiver are probably the real limitation during times of disturbance. If this high-frequency first circuit noise remains constant and the field intensity decreases the 6-A noise will increase in opposite proportion. Therefore it may be assumed that the slopes of the corrected 6-A noise curves in the high noise region will correspond to the slopes of the field intensity distribution curves. In a conservative estimate it is reasonable to assume that this is the case and that although the field intensity falls during times of magnetic disturbance the high-frequency noise will not decrease.

In correcting the less important low noise ends of the 6-A noise curves, use of the field intensity distribution is not so easily justified. It may be reasoned, however, that here atmospheric noise is again low compared to receiver noise but due in this case to a scarcity of electrical storms within favorable transmission distance from the point of reception. Then the corrected 6-A noise distribution at the low noise end would also correspond in shape to the field intensity distribution. For these reasons it is assumed in the absence of better data that the field intensity distribution may be used to correct for the bends that occur at both ends of the 6-A noise distribution curves. Fairly dependable field intensity data are available over a much wider decibel range than is accurately covered by the 6-A noise measurement.

In Fig. 5 is shown by the full line *e–c–d–f* a form of noise-to-signal distribution which is conservatively representative of that experienced on several short-wave radio telephone circuits as described above. The horizontal decibel scale in this figure is arbitrarily referred to the midpoint of the distribution curve. The broken line extension *d–b* represents the decibel distribution of the lowest 15 per cent of the field intensity values as experienced during the years 1930 and 1932. The broken line extension *a–c* similarly represents the distribution of the highest 30 per cent. The reason for using the transatlantic data is that there are many more measurements available in the low field region than there are for transmission over less disturbed paths. The available data indicate that if suitable frequencies are used at all times of the day the distribution of field intensities within the lowest 15 per cent and the highest 30 per cent has roughly the same average slope during different years and on different circuits.

[3] R. K. Potter, "High Frequency Atmospheric Noise," *Proc. I.R.E.*, Vol. 19, pp. 1731–1765, October, 1931.

Circuit Reliability Curve

The corrected form of the noise-to-signal distribution curve represented by the line *a–c–d–b* in Fig. 5 may by a simple translation be put in terms of percentage lost circuit time versus transmission improvement in decibels where noise rather than quality degradation or

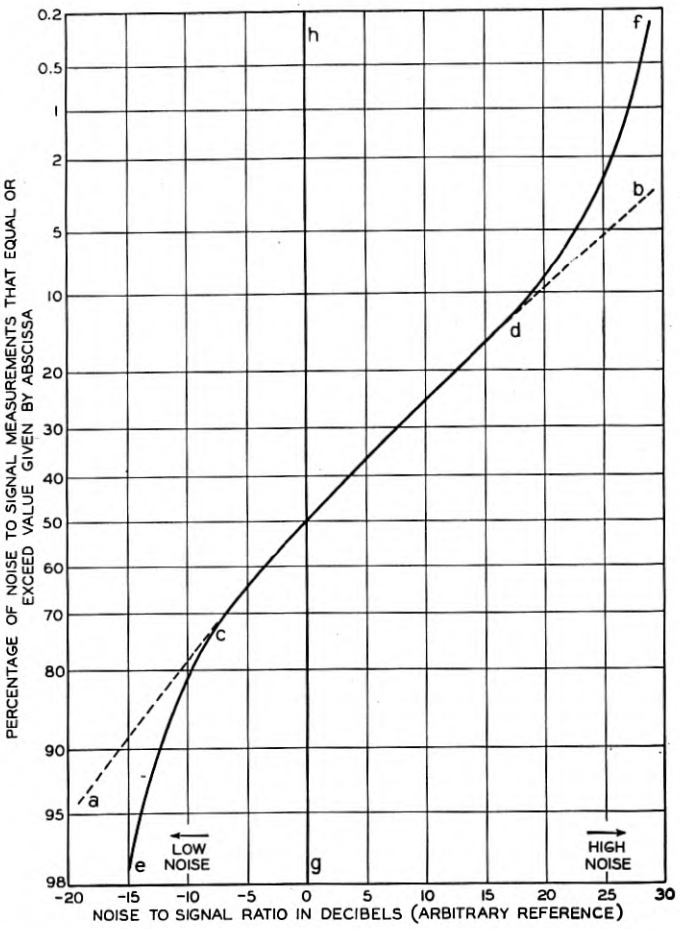


FIG. 5—Correction of assumed 6-A noise distribution curve where it is influenced by terminal equipment effects.

equipment failures interrupts continuity of service. To start this translation, take, for example, the point at which 50 per cent of the noise-to-signal values are greater and 50 per cent are less than a value given on the horizontal scale. This is 0 db and the value in itself has

no significance except as an arbitrary reference point on this scale. A vertical line is erected through this point as shown by *g–h*. If, on a certain short-wave circuit, conditions are unsatisfactory 50 per cent of the time, the effect of 10 db improvement upon this percentage may be determined by shifting the curve *a–c–d–b* of Fig. 5 ten db to the left and reading the percentage value on the vertical scale opposite the intersection of the vertical reference line. It will be remembered from previous discussion that such a shift of the 6-A noise curve toward lower values accompanies a corresponding db transmission improvement. If 50 per cent of the time conditions were unsatisfactory in the former case, they would be unsatisfactory only some 25 per cent of the time for the same tolerable noise condition and 10 db improvement. That is, the lost circuit time has been reduced from 50 to 25 per cent by 10 db transmission improvement.

By shifting the curve *a–c–d–b* of Fig. 5 various amounts to the right and left and tabulating the percentages obtained as described above, a generalized "reliability" curve may be plotted which shows the transmission improvement required to reduce the lost circuit time by any desired amount. Similarly, if we know the percentage lost time on two circuits their transmission performance may be compared on a decibel basis by determining the horizontal db separation between these two lost time values on the "reliability" curve.

A "reliability" curve of the kind described above is shown in Fig. 6. Although it is obviously unsafe to conclude on the basis of the data presented that this curve is accurately representative of all long-range short-wave circuits and circuit conditions, it serves to indicate the order of service improvement that will be afforded within the practical range of transmission improvement. For example, to reduce the lost or unsatisfactory circuit time from 50 per cent to 25 per cent appears to require about 10 db transmission improvement on any long-range short-wave circuit. Starting with a 50 per cent lost time condition and applying improvements in 10 db steps the successive percentages of lost circuit time would be roughly 25, 10, 2.5, 0.7 and 0.1.

Changes in the standards of tolerable service may be treated as equivalent to a change in the effectiveness of transmission as described earlier. Thus in terms of a high grade service the lost circuit time might for example be 50 per cent. For a grade of service 10 db lower than this the lost circuit time would be reduced to 25 per cent. The effect is equivalent to improving the transmission 10 db for the same standard of service.

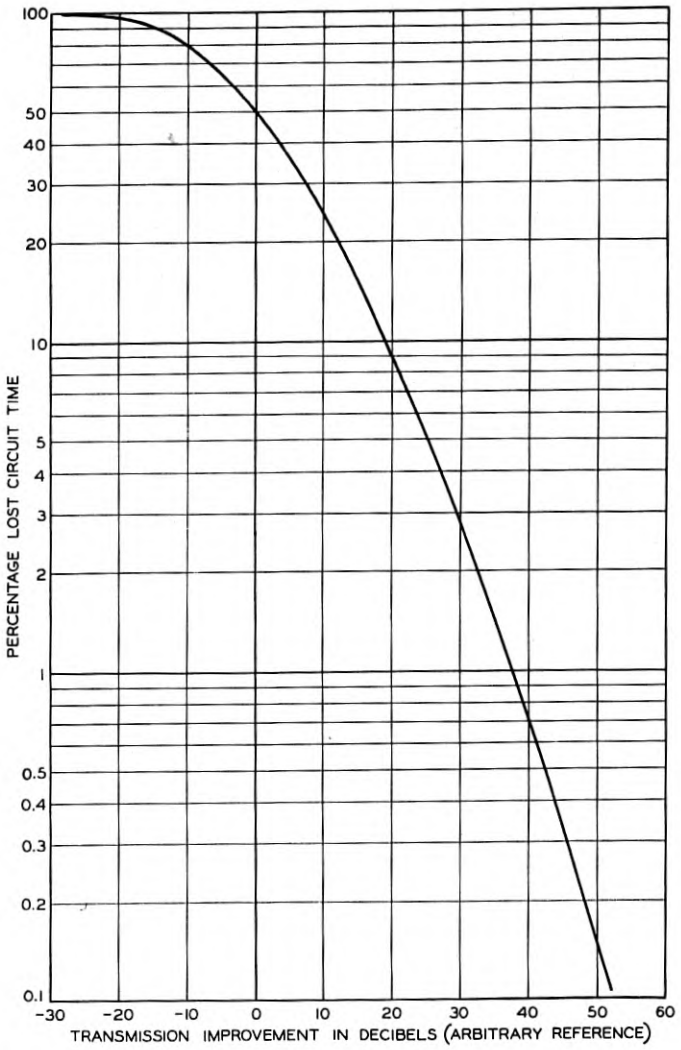


FIG. 6—Approximate relation between transmission improvement and percentage lost circuit time.

Duration of Interruptions

In the traffic operation of radio telephone circuits we are concerned in a practical way with the effect of transmission improvement upon the duration of intervals when circuits are unsatisfactory as well as upon the percentage lost or unsatisfactory circuit time. Obviously, the reduction of lost time must be accompanied by a reduction of the length of unsatisfactory intervals, or what may be termed interruptions to service. It is of interest to know how the distribution of interruptions of various lengths may be expected to vary with improvement of a circuit.

In Fig. 7 the upper curve shows how the number of circuit interruptions for the year as indicated by the ordinate value varied with hours duration on the transatlantic short-wave radio telephone circuits during 1930.[4] This upper curve was obtained from traffic data. In determining the points shown it was necessary to exclude all interruptions of uncertain length that occurred at the beginning or end of the periods when circuits were in use so that only the slope of this curve is significant.

From the field intensity measurements obtained regularly on the transatlantic circuits it is possible to obtain a useful check on the traffic experience. For example, the interruptions may be defined as the intervals of time during which no signal could be heard by beating in the carrier received on a short-wave measuring set to an audible tone with a local oscillation. By this means signals 30 db or more below those required for a barely satisfactory radio telephone circuit can be heard. In Fig. 7 the lower curve shows the distribution of interruptions based upon such a standard. This curve has substantially the same logarithmic slope as the one obtained from traffic experience. That is, the slope remains the same when the conditions defining the point of interruption are shifted by perhaps 30 or 40 db.

The summation of all interruptions shown by curves such as those in Fig. 7 should agree with the observed lost circuit time. If it is assumed, as the rather meager evidence cited above appears to indicate, that the logarithmic distribution of the interruptions remains constant for different circuit conditions, it is possible to show how the probable number of interruptions of various lengths will vary with transmission improvement. With the slope shown the position of curves corresponding to different db improvements is established by the reliability curve of Fig. 6 and the requirement that the summation of interruptions equal the lost time. Curves obtained in this manner are shown

[4] Interruptions during other years have been too infrequent to provide dependable data.

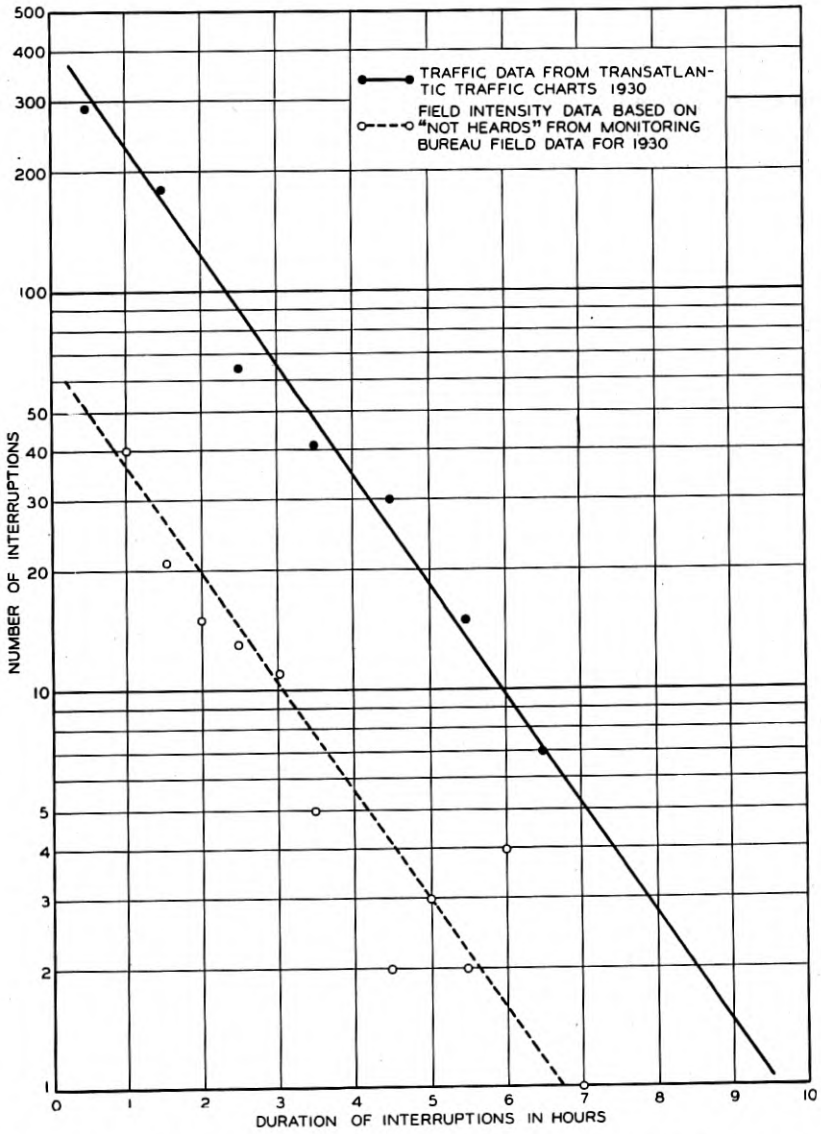


FIG. 7—Observed relation between number and duration of transatlantic short-wave radio telephone circuit interruptions during part of 1930.

in Fig. 8. As for the generalized reliability curve of Fig. 6, the reference point adopted is 50 per cent lost circuit time. Consequently the curve of Fig. 8 marked "50 per cent lost circuit time" is also designated as "0 db Transmission Improvement." Knowing the per-

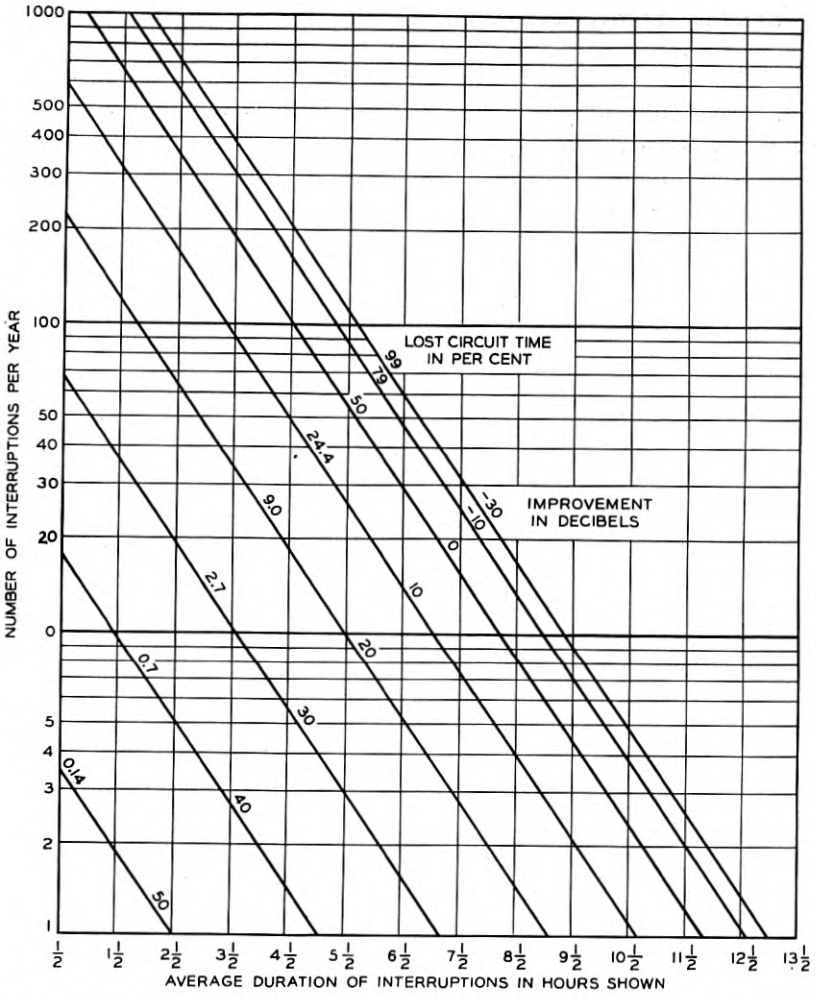


FIG. 8—An estimate of probable duration versus number of interruptions for full time operation.

centage lost time experienced on a certain circuit the position of the corresponding interruption curve in Fig. 8 may be determined by interpolation. The effect of, say, 10 db transmission improvement

upon the probable interruptions, is determined by shifting this curve 10 db to the left as measured by the indicated db spacing between curves.

It should be remembered that estimates based upon Fig. 8 are probably very approximate but the curves will at least serve to indicate the trend of improvement effects.

ACKNOWLEDGEMENT

Acknowledgement is due the various engineers and operators of the British General Post Office and of the American Telephone and Telegraph Company who were responsible for the accumulation of the data used in this analysis.

Spontaneous Resistance Fluctuations in Carbon Microphones and Other Granular Resistances

By C. J. CHRISTENSEN and G. L. PEARSON

Voltage fluctuations which occur in resistance elements of the granular type when a direct current is flowing have been measured in the granular carbon microphone, commercial grid leaks, and sputtered or evaporated metal films. The results can be experessed by the formula

$$\overline{V_c^2} = KV^\alpha R^\beta \log (F_2/F_1),$$

where $\overline{V_c^2}$ is the mean square fluctuation voltage, V is the d.-c. voltage across the resistance R, α and β are constants having values of about 1.85 and 1.25, respectively, and F_2 and F_1 are the limits of the frequency range over which the fluctuation voltage is measured. The constant K depends, among other things on the temperature, the surrounding medium, and the dimensions and material of the resistance element; for a commonly used carbon transmitter at ordinary operating conditions its value is about 1.3×10^{-11}.

The spontaneous voltage fluctuations and the signal due to acoustic modulation are affected in almost an equivalent manner by the applied d.-c. voltage which suggests that the two effects arise from the same type of mechanism, namely a fluctuating resistance at the points of contact between granules. Experiment shows that although the acoustic signal produces a resistance modulation which is in phase at all contacts the spontaneous resistance fluctuations are completely random.

On the assumption that a region of secondary conduction, wherein the resistance fluctuation lies, surrounds each area of primary conduction as postulated in recent contact theory a value of β consistent with experiment has been deduced. On the further assumptions that thermal energy produces the mechanical fluctuations and that the equipartition law governs the distribution of energy between oscillators the observed frequency distribution follows.

Introduction

WHEN a direct current is passed through certain types of resistance elements a small potential fluctuation between the terminals of the resistance can be observed in addition to that caused by the thermal agitation of electric charge. The resistances in which this effect is particularly noted are granular carbon microphones and commercial grid leaks which are granular in nature, such as sputtered or evaporated metal films, and any of a number of composite materials containing carbon in a finely divided state. If such a resistance element is in a current-carrying circuit associated with a telephone receiver or loud speaker, particularly when amplification is present, a steady hissing noise which sounds like that due to shot effect or thermal agitation of electric charge is heard. It is this noise which sets a practical limit to the use of the carbon microphone in sound fields of low intensity, and of commercial grid leaks in circuits carrying direct current and working at low signal levels.

The resistance of a granular conductor has been shown experimentally to lie almost entirely within very small volume elements in the regions of the contact areas.[1] It is our hypothesis that there exist minute fluctuations of resistance in the region of contact, and when such an element carries direct current a potential fluctuation between the terminals can be observed. Accordingly we propose for this phenomenon the term "contact noise."

In a study of the electrical disturbances in a carbon transmitter Kawamoto [2] found that in addition to "carbon burning," which is a sharp crackling noise sometimes present in the carbon transmitter when the voltage across individual contacts is of the order of 0.5 volt or greater,[3] there is a continuous rushing sound which is always present no matter how well the transmitter is shielded from external disturbances. Kawamoto applied the term "carbon roar" to this phenomenon. Frederick [4] in discussing the disturbances in the carbon transmitter states that the noise power is proportional to the square of the direct current passing through the transmitter. More recently Otto [5]—who has been working on this subject contemporaneously with ourselves—has reported the results of an extended investigation of this phenomenon. The present report parallels to some extent the study of Otto but in addition new aspects of the phenomenon have been investigated, more accurate data have been obtained, and the conclusions drawn from these experimental results are fundamentally different from those of Otto.

Electrical disturbance in grid leaks, which becomes evident with the passage of current, was first reported by Hull and Williams [6] who observed the phenomenon in resistances formed by an India ink line. Preliminary reports have since been published concerning such noise in thin metallic films on glass.[7] The observations of Otto [5] were also extended to fine carbon wires and copper-oxide resistances. More recently Meyer and Thiede [8] have investigated the noise in resistances consisting of thin films of carbon on a refractory base.

We have performed noise measurements on each of the types of resistance elements mentioned above and the experimental results

[1] F. S. Goucher, *Jour. Franklin Inst.* **217**, 407 (1934); *Bell Sys. Tech. Jour.* **13**, 163 (1934).

[2] T. S. Kawamoto, Unpublished Report, Engineering Division, Western Electric Company, April, 1919.

[3] This disturbance undoubtedly has its origin in the heat generated at the carbon contact by the passage of current.

[4] H. A. Frederick, *Bell Telephone Quarterly* **10**, 164, July, 1931.

[5] R. Otto, *Hochfrequenztechnik und Elektroakustik* **45**, 187 (1935).

[6] A. W. Hull and N. H. Williams, *Phys. Rev.* **25**, 173 (1925).

[7] G. W. Barnes, *Jour. Franklin Inst.* **219**, 100 (1935).

[8] Erwin Meyer and Heinz Thiede, *E.N.T.* **12**, 237 (1935).

which are presented in this paper indicate that the noise observed in each case is of the same nature and is traceable to the existence of contacts between granules or perhaps granular boundaries.

Apparatus

The experimental arrangement used in the measurements to be described here is given in schematic form in Fig. 1. The system includes the input circuit, a high gain amplifier, appropriate filters, attenuator and output measuring device.

The input circuit consists of the resistance under test, a battery for supplying the direct current, a potentiometer for measuring resistance and voltage, a standard signal oscillator for calibration purposes and appropriate resistances and condensers for coupling to the amplifier. In some cases an input transformer having a high-turns ratio was also required in order to raise the signal level above the amplifier noise level. The granular resistance element was shielded from acoustical, mechanical and electrical shocks by suspending it with rubber bands

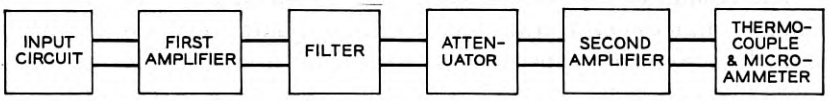


Fig. 1—Schematic amplifier circuit for measuring contact noise in granular resistance elements.

inside a tightly sealed iron box which was lined with alternate layers of hair felt and ¼-inch sheet lead. The remaining parts of the input circuit were also carefully shielded.

The high-gain amplifier consists of two separate resistance coupled units, each containing three stages. Each unit is so designed and shielded that the effect of external disturbances is eliminated. The total gain obtainable is about 165 db, with the frequency response uniform to within 2 db from 10 cycles to 15,000 cycles. In most of the measurements described here, however, a filter which transmitted only those frequencies above 100 cycles was inserted between the first amplifier unit and the attenuating network.

The gain of the amplifying system could be varied in steps of 20 db by means of interstage potentiometers. In addition, a 600-ohm attenuator having a range of 63 db in steps of 1 db was placed between the filter circuit and the second amplifier unit. The output measuring instrument was a 600-ohm vacuum thermocouple and microammeter. The deflection of the meter was closely proportional to the mean square voltage applied to the couple. Individual noise measurements were

made by adjusting the attenuation so as to bring the deflection of the microammeter as near mid-scale as possible. Fractions of a db were estimated by means of the deviation from the standard mid-scale reading. Thus what was measured in each case is the insertion loss necessary to produce a standard electrical output.

Noise as a Function of Applied d.-c. Voltage

The contact noise in several different types of granular resistance elements was measured as a function of the applied d.-c. voltage, all other variables such as resistance, frequency range, temperature, etc., being held constant. The first of these measurements to be described is that obtained by using a standard handset telephone transmitter. The circuit used for coupling to the high-gain amplifier is shown in the insert of Fig. 2, the essential parts being an input transformer having a high-turns ratio, a d.-c. voltage supply, and a standard a.-c. signal generator. The resistance of the carbon transmitter was about 50 ohms.

The results of the measurement are shown in Fig. 2 where mean square contact noise voltage is plotted as ordinate and the d.-c. voltage directly across the transmitter is plotted as abscissa, the scale being logarithmic in each case. Measurements were made as the transmitter voltage was varied from 0.00145 to 4.5 volts. This is the greatest possible voltage range in which contact noise can be observed in this instrument since the contact noise is masked at the higher voltages by carbon burning and at the lower voltages by the thermal noise of the transmitter resistance. Thus the total noise at 0.00145 volt is only slightly above thermal noise and the measured value consists of thermal plus contact noise. The two effects have been calculated separately and the latter plotted as a cross. Using this method of plotting it is seen that there is a straight line relationship between contact noise and voltage over the entire lower range. These experimental data can be accurately represented by the equation

$$\overline{V_c^2} = \text{Const. } V^{\alpha}, \tag{1}$$

where $\overline{V_c^2}$ is the mean square contact noise voltage, V is the d.-c. voltage across the transmitter, and α is a numerical constant having in this case the value 1.85.

By this procedure the contact noise in a number of types of carbon transmitters, filled with carbons of various origins, was measured as a function of voltage. In each case the relationship given by Eq. (1) was followed very closely over a wide range of voltages. The value of

α varied slightly from cell to cell, the extreme values being 1.75 and 1.97 with an average of about 1.85.

Figure 3 gives the results of alternate measurements of contact noise and acoustic modulation performed on a particular telephone trans-

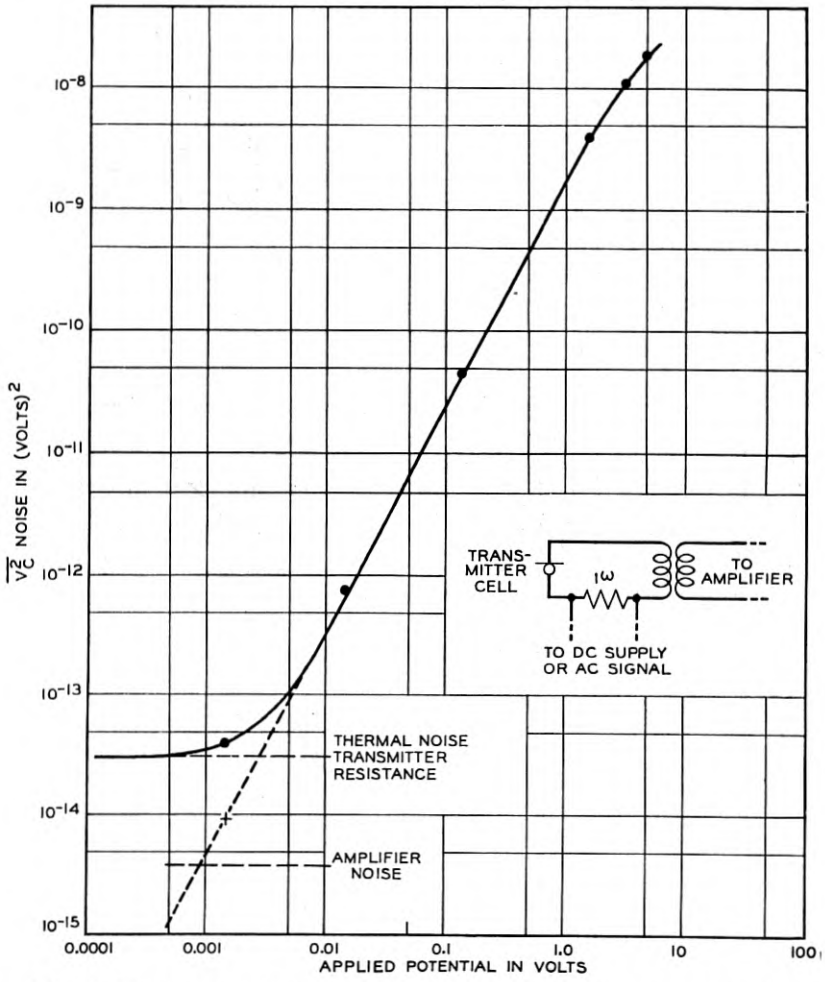


Fig. 2—The mean square contact noise voltage in a standard carbon transmitter as a function of the d.-c. voltage on the instrument. The noise level of the amplifier and the thermal noise level of the transmitter are indicated.

mitter. The acoustic field was of constant frequency and was supplied by an accurately controlled oscillator and "artificial mouth." The sound field was of such intensity that when the transmitter output was measured the background noise gave only an inappreciable part of the

whole energy. In this figure the abscissæ represent the d.-c. voltage directly across the transmitter, the scale being logarithmic, and the ordinates represent mean square contact noise or acoustic signal voltage plotted in db above an arbitrary zero level. The experimental plots for both signal and noise are straight lines but of slightly different slope.

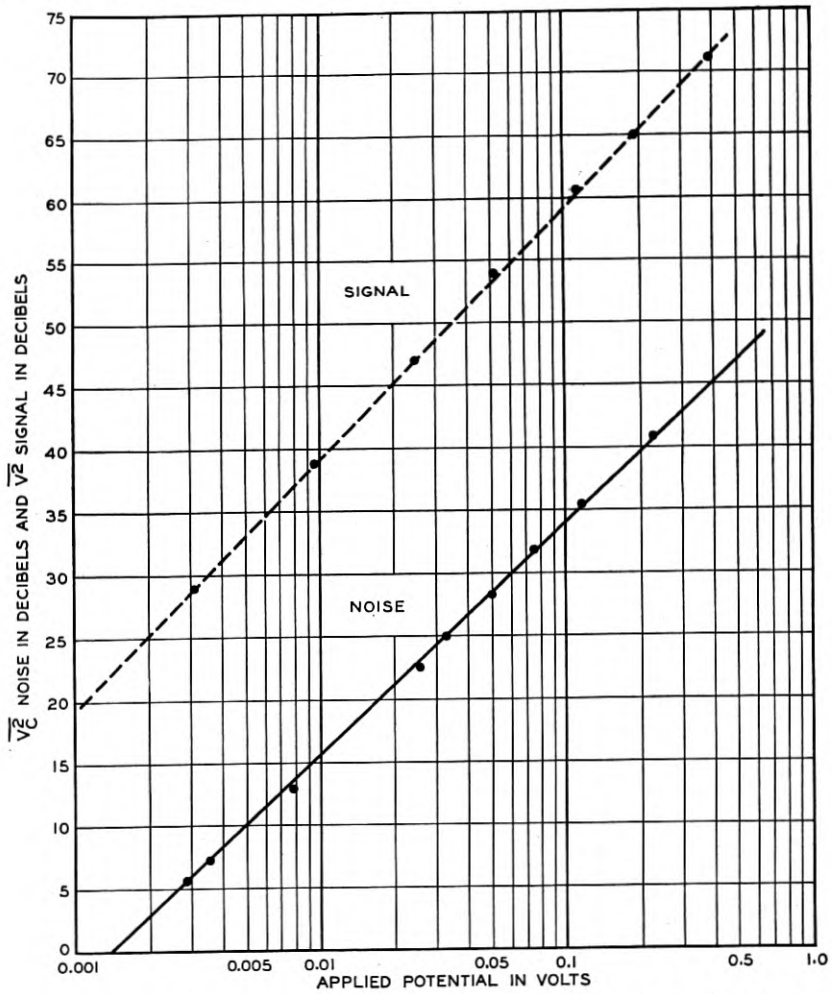


Fig. 3—The mean square contact noise and signal voltages in a standard carbon transmitter as a function of the d.-c. voltage on the instrument. The signal was obtained from a constant intensity sound field.

Analysis of the data shows that the mean square signal voltage due to acoustic modulation is accurately proportional to d.-c. transmitter voltage squared while the noise curve fixes the value of α at about 1.85. In the case of the signal the square relationship is to be expected since

the sound field produces a constant resistance modulation at the points of contact between carbon granules and the granular aggregate obeys ohms law over the entire voltage range. The fact that the noise and the signal follow so nearly the same relationship indicates that the noise also arises from resistance modulation at the points of contact between carbon granules. Since α is slightly less than 2, however, the noise mechanism is not entirely independent of the applied voltage.

Noise as a function of applied voltage was also measured in single contacts between carbon particles. For these observations a cantilever bar device was used in which the contact can be rigidly fixed and manipulated at will. This apparatus, shown in Fig. 4, consists of a

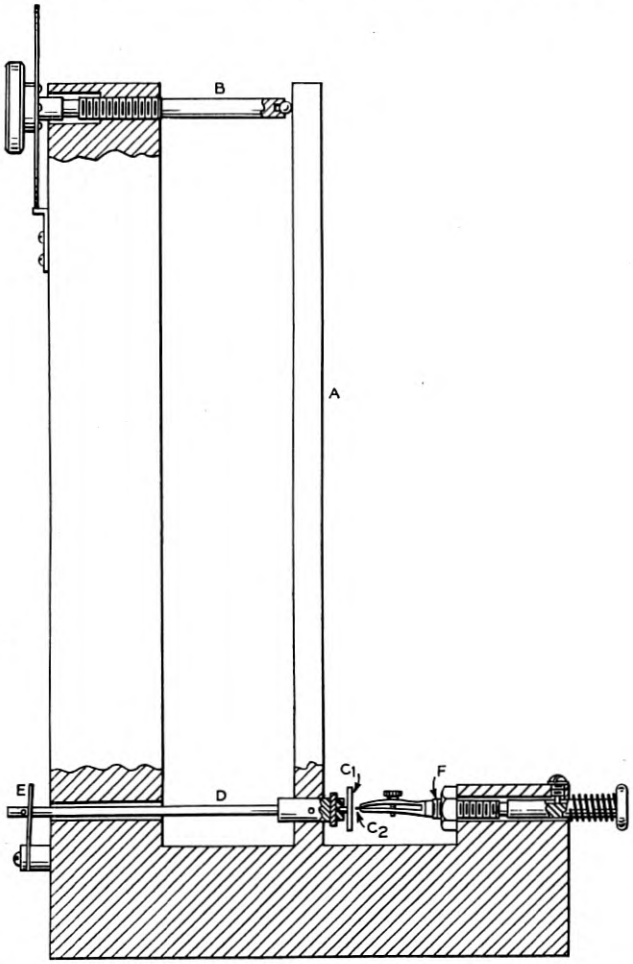


Fig. 4—Diagram of cantilever bar device for producing small contact displacements.

cantilever bar A integral with a massive L-shaped base, the entire device having been milled from a single piece of steel. A given displacement of the graduated screw B produces a greatly diminished displacement of the movable electrode C_1. The dimensions of the bar were so chosen that contact displacements of the order of 1×10^{-7} cm. could be produced. The motion of C_1 is made strictly linear by means of the pivoted rod D and any slack motion is eliminated by the

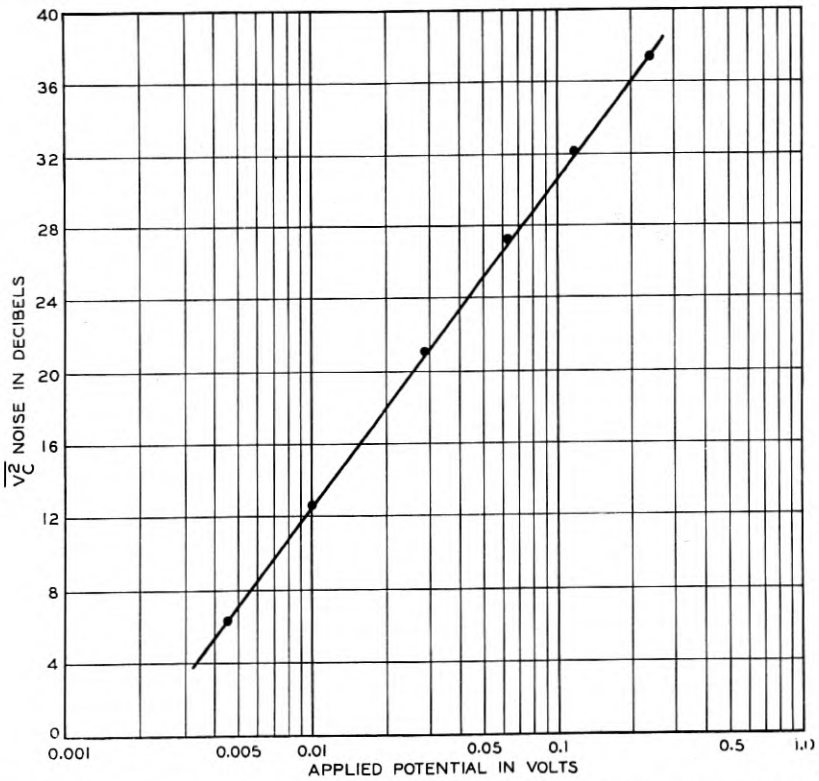


Fig. 5—The mean square contact noise voltage in a single carbon contact as a function of the applied voltage. The contact resistance was held constant at 76 ohms.

spring E. The contact is initially adjusted by the graduated screw F on which the stationary electrode C_2 is mounted. C_1 consisted of a flat polished disk of carbon like that used in the desk set transmitter, while C_2 was a composite carbon spheroid clamped securely between two gold plated jaws. Both the carbon plate and the spheroid were coated with a pyrolitic deposit of hard carbon.

The noise in a large number of single carbon contacts, connected in place of the transmitter in the input circuit shown in Fig. 2, was

measured as a function of the voltage on the contact, the resistance being held fixed. In every case the general law given by Eq. (1) was found valid, α varying between the limits 1.75 and 1.95 for different contacts. The results of a typical measurement on a contact having a resistance of 76 ohms are shown in Fig. 5. The experimental points fall on a straight line having a slope corresponding to a value of α equal to 1.83.

Figure 6 gives the results of contact noise measurements performed on a commercial grid leak which was made by coating a thin layer of

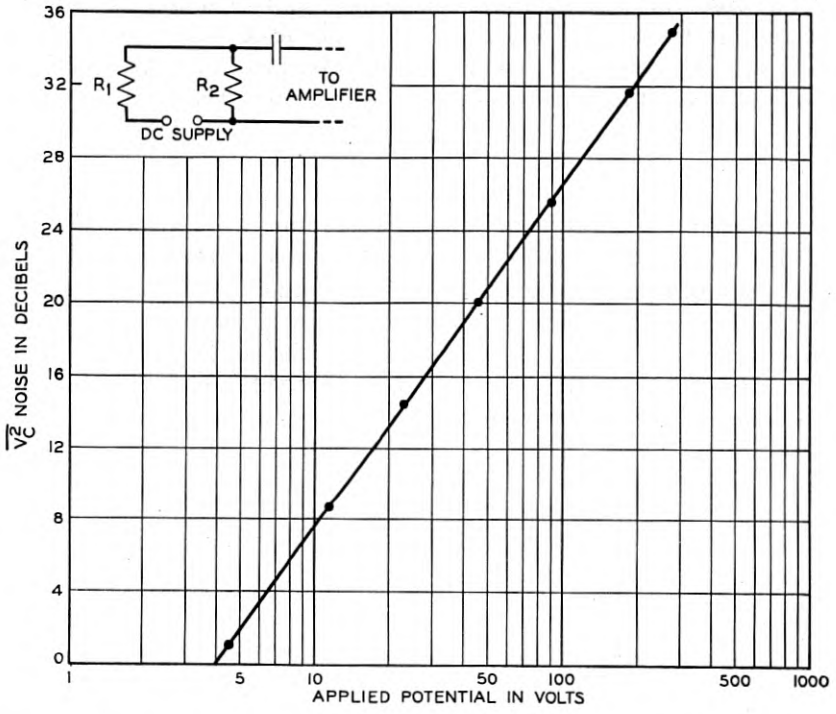


Fig. 6—The mean square contact noise voltage in a 50,000-ohm carbon grid leak resistor as a function of the applied voltage.

finely divided carbon and binder on glass. The input circuit shown in the insert, consists of R_1, the sample under test; R_2, a metal wire resistor which produces no contact noise; and a suitable source of d.-c. voltage. The resistance of both R_1 and R_2 was 50,000 ohms. The experimental points lie on a straight line the slope of which fixes the value of α at 1.90. Individual points could be reproduced within an

accuracy of 0.1 db. Similar measurements of noise in thin metallic films deposited by either the cathode sputtering or evaporation process gave results in agreement with Eq. (1), the value of α lying between the limits mentioned above.

Thus it is seen that all types of granular resistance elements which we have tested, namely, carbon transmitters, single carbon contacts and those consisting of thin films of carbon or metal follow the same relationship for noise as a function of applied d.-c. voltage.

Noise as a Function of Contact Resistance

The observation of contact noise as affected by contact resistance is necessarily limited to loose contacts, for in fixed resistance elements such as grid leaks and conducting films one has no means of independently varying their resistances. One alters the resistance of a single contact by the relative displacement of the two contacting elements. The resistance of a multi-contact device, such as a carbon microphone, may be altered either by a relative displacement of the contacting particles, or by a change in the number of contacts between the electrodes. The noise is affected differently by these two methods of resistance change; hence one must study them separately. In this section we shall be concerned only with noise as affected by resistance changes due to contact displacement both in single contacts and in aggregates; that due to a change in the number of contacts between the electrodes will be considered in our discussion of the noise from a contact assemblage.

Figure 7 is a diagram of the input circuit used to study the relation-

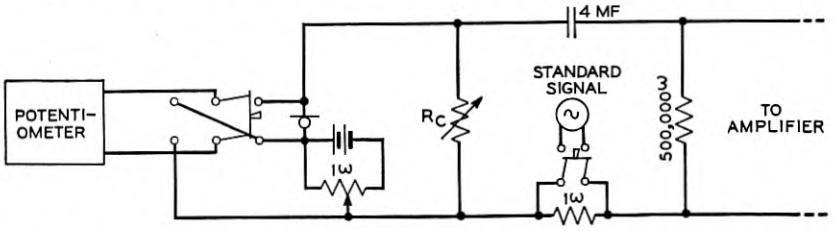


Fig. 7—Diagram of the circuit used in measuring contact noise as a function of resistance in granular resistance elements.

ship between noise and resistance. By means of the potentiometer we could measure both the voltage supplied to the contact circuit and that across the contact or transmitter. The resistance R_c is so adjusted for each noise observation that one half the voltage supplied to the circuit is across the contacts. The contact resistance for this condition is given by R_c. Also, in every case, one half the generated noise voltage is impressed on the input tube of the amplifier.

The single contacts studied were mounted in the cantilever bar device as described in the preceding section. We found it important to wait after the mounting of a contact long enough for the whole bar to come to thermal equilibrium before a measurement was attempted, otherwise very erratic results were obtained. Figure 8 is a typical curve obtained when the mean square noise voltage is plotted in db against the contact resistance on a logarithmic scale. For this

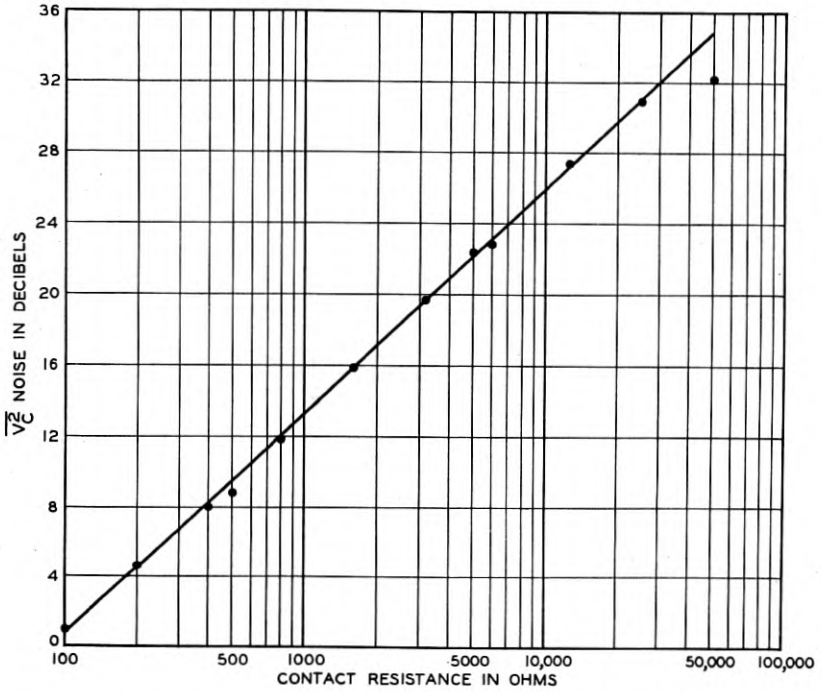


Fig. 8—The mean square contact noise voltage in a single carbon contact as a function of the contact resistance. The resistance was varied by changing the contact displacement while the applied voltage was held at 0.1 volt.

measurement the d.-c. voltage on the contact was 0.1 volt. The experimental relationship between noise and resistance is given by

$$\overline{V_c^2} = \text{Const. } R^{\beta}. \tag{2}$$

The data plotted in Fig. 8 give the value $\beta = 1.25$, which is the average value found for all the contacts studied. For individual contacts β varied between the extremes of 1.1 and 1.42. The studies on other properties of single contacts also exhibit a rather wide variability from contact to contact, hence the above result is not surprising.

A departure from the straight line relationship, such as is plotted in Fig. 8, occurs only when the contact is in a relatively high-resistance state due to a very slight compression of the contacting particles. When in this condition contacts are quite unstable and observations upon them erratic, but, in general, the noise originating in them is less than that expected if Eq. (2) held over the entire range of contact resistance values.

For the study of an aggregate of contacts the carbon cell from a barrier type transmitter was chosen. The structure of this cell (see insert Fig. 9) is such that one would expect the major portion of the

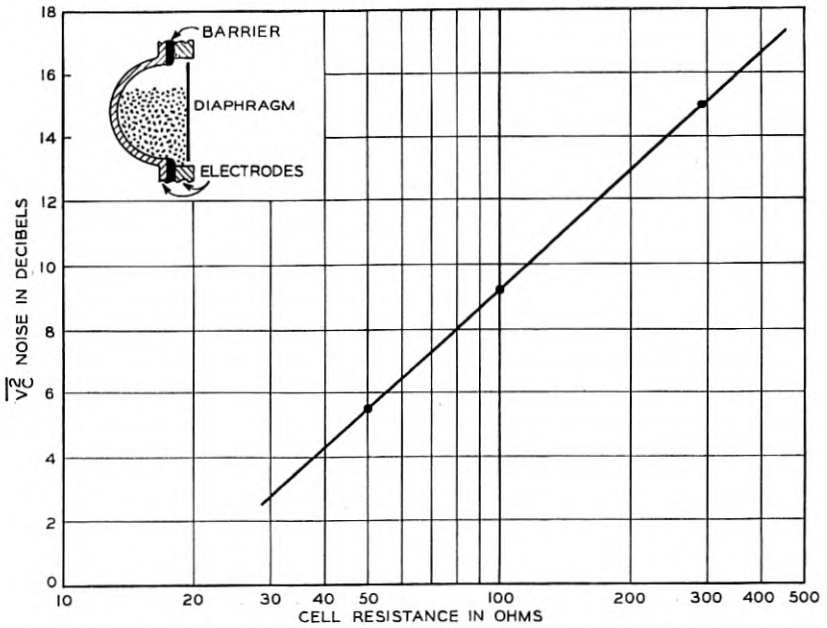


Fig. 9—The mean square contact noise voltage in a standard carbon transmitter as a function of the cell resistance. The resistance was varied by changing the amount of carbon filling while the transmitter voltage was held constant at 1.0 volt. Each experimental point represents the average of nine different readings.

current to be conducted through a small part of the total mass of granules near the bottom of the cell, and carbon added to the top of the cell to act mainly to increase the contact compressions of the conducting contacts. The electrodes of this cell are heavily gold plated, which assures that the resistance observed is almost entirely that of the contacts within the aggregate, a condition which is not fulfilled when carbon electrodes are used.

Figure 9 summarizes the results obtained in a typical set of measurements on the noise generated in the transmitter cell before mentioned. The resistance was varied by changing the height of the carbon layer above the carbon which was in the conducting path. Each of the plotted points is the average of nine observations, all of which occur in a range of 1 db. The relationship plotted in Fig. 9 can also be expressed by Eq. (2) and we again find $\beta = 1.25$. We shall see later—Eq. (8)—that when the resistance of an aggregate is altered by changing the number of conducting contacts between the electrodes quite another relationship between noise and resistance is obtained. Hence our assumption regarding the nature of the resistance change in this cell is consistent with the data of Fig. 9, and we believe that in this experiment we have measured the average value of β for all the contacts in the conducting path and have found it to be in agreement with the average value deduced from our single-contact measurements.

Noise as a Function of Frequency

For measuring the frequency distribution of the noise the filter shown in Fig. 1 was replaced by a frequency analyzer [9] having a constant band width of 20 cycles, the midpoint of which could be set at any point between 50 and 10,000 cycles per second. The calibration of the apparatus was checked by measuring the frequency distribution of thermal noise which was constant over this entire range, in accordance with theory.

The results of the measurements on a standard carbon transmitter, maintained at constant resistance and applied voltage, are shown in Fig. 10 where ordinates represent the mean square noise voltage over the 20-cycle band and abscissæ represent the mid-frequency of the band. It is seen that the experimental points fall on a straight line having a negative slope of about 1.0. This relationship may be represented by the equation

$$\Delta \overline{V_c^2} = \text{Const. } \Delta F/F, \tag{3}$$

where $\Delta \overline{V_c^2}$ is the mean square noise voltage for the frequency band ΔF. Integrating Eq. (3) between fixed limits we obtain:

$$\overline{V_c^2} = \text{Const. } \log (F_2/F_1), \tag{4}$$

which gives the total noise over the frequency range F_1 to F_2.

Figure 11 gives the results of similar measurements on a high-

[9] T. G. Castner, *Bell Laboratories Record* **13**, 267 (1935).

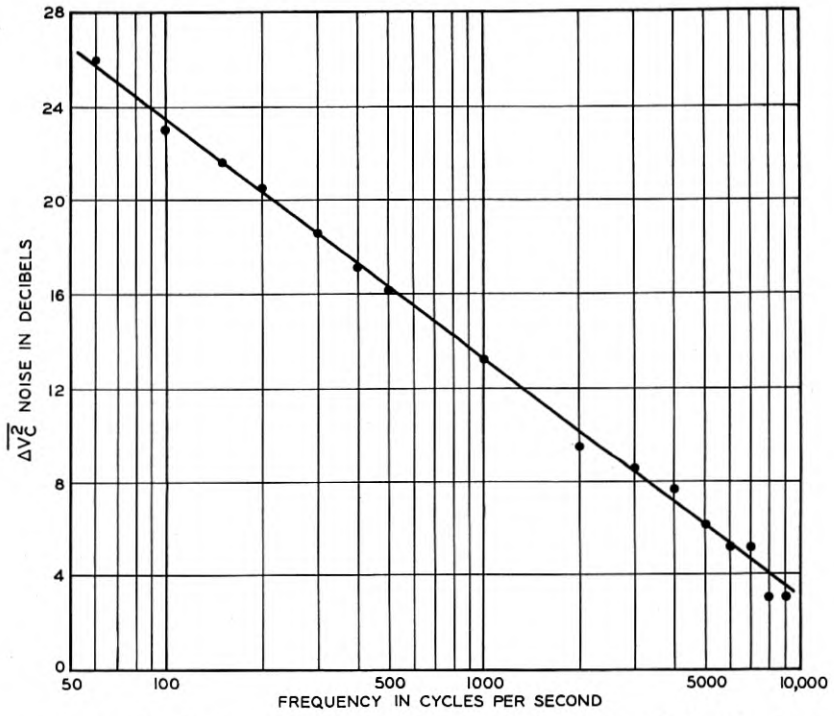


Fig. 10—The frequency distribution of contact noise in a standard carbon transmitter for normal operating conditions. The ordinates give the mean square contact noise voltage in a twenty-cycle band the midpoints of which are given by the abscissæ.

resistance carbon grid leak. It is seen that the noise has precisely the same frequency distribution in both the carbon transmitter and in the grid leak, which further supports our belief that the noise mechanism is the same in each case.

Otto [5] reported similar measurements of noise as a function of frequency in carbon transmitters, single contacts of carbon, carbon grid leaks and copper oxide resistances. Whereas we find an almost exact inverse relationship between noise and frequency for all types of elements tested he shows curves with negative slopes ranging from 1.0 to 1.4. Meyer and Thiede [8] in measurements on thin carbon films obtained negative slopes having values between 1.0 and 2.0.

Contact Noise as a Function of Temperature and Surrounding Medium

For the complete elucidation of contact noise the knowledge of its dependence upon temperature is important. However, the difficulties involved in such a measurement are so great that we have been unable,

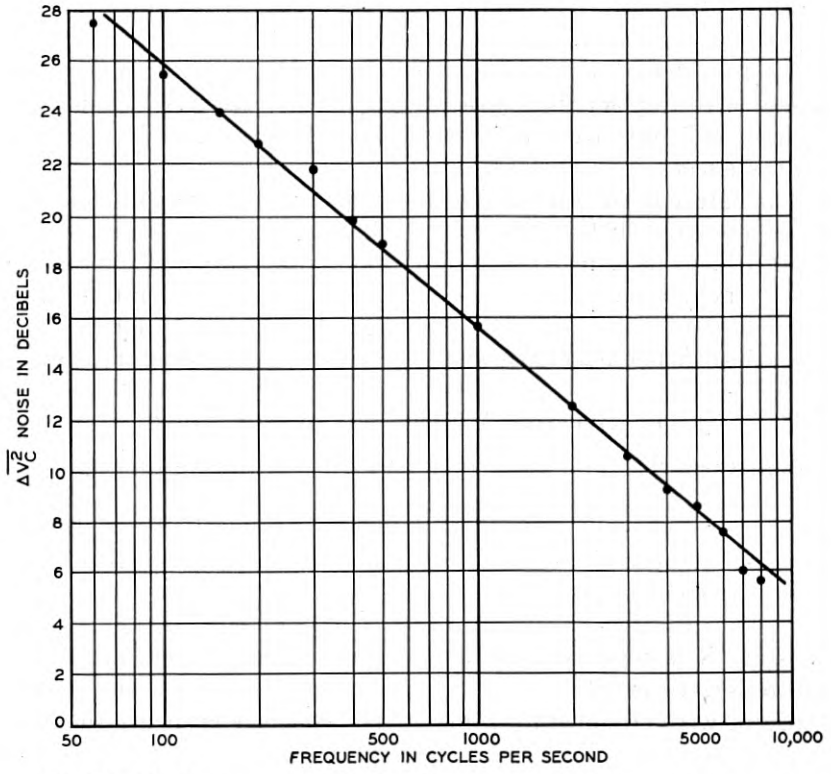


Fig. 11—The frequency distribution of the contact noise in a 50,000-ohm carbon grid leak resistor. The method of plotting the experimental points is the same as that shown in Fig. 10.

as yet, to obtain dependable results. For a satisfactory observation of the effect of temperature on the noise of contacts one must be certain that the conducting area in a contact remains constant and independent of temperature. Temperature variations can alter the conducting area in at least two ways; the contacting particles may be relatively displaced due to differential thermal expansions of the apparatus, and the change of the quantity of adsorbed gas on the contacting surfaces can alter the contacting areas without any relative displacement of the contacting particles. Both of these conditions are very difficult to control in any measurement involving temperature changes. However, our measurements of the relationship between contact noise and temperature, performed under the most carefully controlled conditions which we have been able to apply, indicate that contact noise may change either positively or negatively as a function

of temperature depending upon the conditions of the experiment, but that the total change is not more than 3 or 4 db in the temperature range from 90 to 300 degrees Kelvin. These observations are consistent with the fact that Otto [5] has reported a decrease of noise with increase of temperature while Meyer and Thiede [8] have reported the reverse effect.

The nature of the surrounding medium seems to affect the intensity of the noise but little. The noise of the contact under oil seems to be slightly less, one or two decibels, than when the contact is in a vacuum of 10^{-5} mm. of mercury. The noise in air seems to be intermediate between these two extremes. This leads us to believe that the noise mechanism is not associated with the medium surrounding the contact.

Quantitative Values of Contact Noise

Equations (1), (2) and (4) may be combined to give the expression

$$\overline{V_c^2} = K\, V^{\alpha} R^{\beta} \log (F_2/F_1). \tag{5}$$

This is the general empirical equation found for noise in granular resistance elements as a function of voltage, resistance, and frequency. The average values for α and β are respectively 1.85 and 1.25. The constant K is dependent on the material, shape, temperature, etc. of the resistance element. The following representative values of this constant were obtained for some of the resistance elements which we have measured:

Single carbon contact	1.2×10^{-10}
Western Electric No. 395-B telephone transmitter	1.3×10^{-11}
100,000 ohm carbon grid leak	1.1×10^{-21}

For different single carbon contacts this constant did not vary more than 20 per cent as long as a given type of carbon was used, and a change in the type of microphonic carbon produced a variation by not more than a factor of two.

The contact noise in a Western Electric No. 395–B telephone transmitter under actual working conditions (R = 45 ohms, V = 2.5 volts, F_1 = 200 c.p.s. and F_2 = 3000 c.p.s.) is given by Eq. (5) as 9.8×10^{-5} volts. The output signal of this transmitter for standard voice operation is about 0.1 volt. The spread between signal and contact noise is so great that this noise is not a disturbing factor in the standard carbon transmitter as used in telephone service. This is not true, however, in the case of high quality carbon transmitters used for studio work and public address systems. The sound fields under the conditions wherein such instruments are apt to be used are much less intense

than for the telephone transmitter under its normal condition of use, and hence, the contact noise becomes a limiting factor when the carbon microphone is used in weak sound fields.

Equation (2) is not suited for representing contact noise as a function of resistance in grid leaks since a change in resistance is brought about by variations in the dimensions and materials of the conducting film rather than by a change in contact displacement as is done in the case of loose contacts. For this reason the constant given above applies only to 100,000-ohm resistances of a given type.

It is of interest to note that the constant for the carbon grid leak resistor is smaller than that for the single contact by a factor of 10^9. This is due, in part, to the fact that the total voltage V across the grid leak resistor is divided among a network of contacts each of which produces noise independently of the others. The total noise from such an assemblage, as will be shown in the following section, is less than that arising from a single contact. This suggests that the contact noise in a solid carbon filament, if it exists at all, should be still smaller than that in the grid leak. We have made measurements on such a filament having a diameter of 0.0025 cm. and a resistance of 75,000 ohms. After taking great precautions to eliminate all the noise at the terminal connections we were unable to detect any noise in addition to that of thermal agitation for d.-c. loads as great as the filament would carry without being destroyed (a current density of 3×10^3 amperes per square cm.).

Noise From a Contact Assemblage

A transmitter cell contains an assemblage of contacts and we have shown that the noise from such an assemblage follows the empirical law set forth in Eq. (5), which also holds for single contacts. Several important deductions are possible when we study the noise from an assemblage as a function of the number and arrangement of the contacts within it.

Assemblage With Contacts in Parallel

Consider n contacts, R_1, $R_2 \cdots R_n$, placed in parallel across a direct current supply and inductance as shown in Fig. 12*A*. The inductance is large enough so that it offers an effectively infinite impedance to the fluctuation voltage we expect to study. Due to the fluctuation of resistance in the contacts and the passage of direct current they will act as a.-c. generators. Figure 12*B* is the equivalent a.-c. circuit, where e_1, $e_2 \cdots e_n$ are the instantaneous a.-c. voltages generated because of the fluctuating resistance in the respective contacts. The instantane-

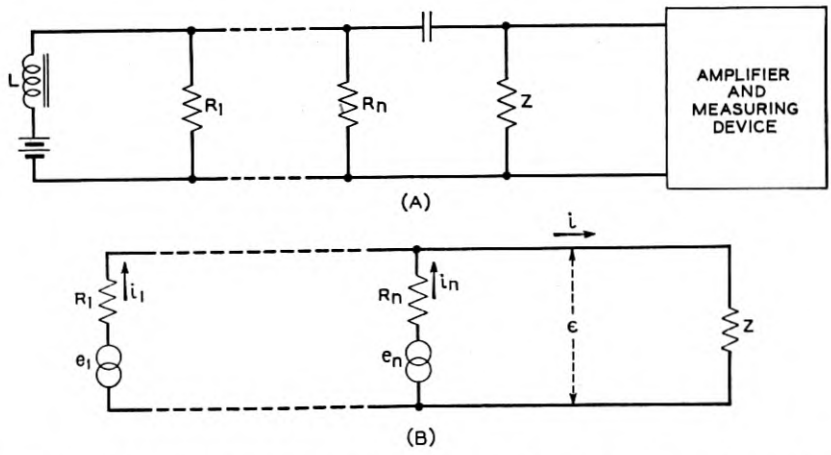


Fig. 12—(*A*) Circuit for measuring the contact noise of a parallel assemblage of resistance elements. (*B*) The equivalent a.-c. circuit of (*A*).

ous value of a.-c. voltage experienced across the impedance Z—which is the input impedance of a measuring circuit—is

$$\epsilon = e_1 - i_1 R_1 = \cdots e_n - i_n R_n = iZ,$$

where $i_1, i_2 \cdots i_n$ are the respective fluctuating currents flowing because of the generator action of the fluctuating contact resistances. Also we have

$$i = i_1 + i_2 \cdots i_n.$$

From these two expressions we get

$$\frac{e_1}{R_1} + \frac{e_2}{R_2} + \cdots \frac{e_n}{R_n} = \epsilon \left[\frac{1}{R_1} + \frac{1}{R_2} + \cdots \frac{1}{R_n} + \frac{1}{Z} \right] = \frac{\epsilon}{Y},$$

where

$$\frac{1}{Y} = \frac{1}{R_1} + \cdots \frac{1}{R_n} + \frac{1}{Z}.$$

Hence we obtain

$$\epsilon = e_1 \frac{Y}{R_1} + \cdots e_n \frac{Y}{R_n}.$$

The noise power dissipated in Z, and thus measured by the measuring device, is defined in the usual way as $\overline{\epsilon^2}/Z$, where $\overline{\epsilon^2}$ is the mean square voltage across the impedance Z and is defined as

$$\overline{\epsilon^2} = \operatorname*{limit}_{t \to \infty} (1/t) \int_0^t \epsilon^2 dt.$$

If the value of ϵ as given above is substituted in this equation, and the relative phases of the individual contact voltages $e_1 \cdots e_n$ are considered random, one derives the relationship

$$\overline{\epsilon^2}_{\text{parallel}} = \frac{\dfrac{\overline{e_1^2}}{R_1^2} + \dfrac{\overline{e_2^2}}{R_2^2} + \cdots \dfrac{\overline{e_n^2}}{R_n^2}}{\left[\dfrac{1}{R_1} + \dfrac{1}{R_2} + \cdots \dfrac{1}{R_1} + \dfrac{1}{Z}\right]^2}. \tag{6}$$

The experimental test of this derived relationship will be given later.

Assemblage with Contacts in Series

Consider n contacts placed in series and supplied with current from a battery through an inductance as shown in Fig. 13*A*. The equivalent

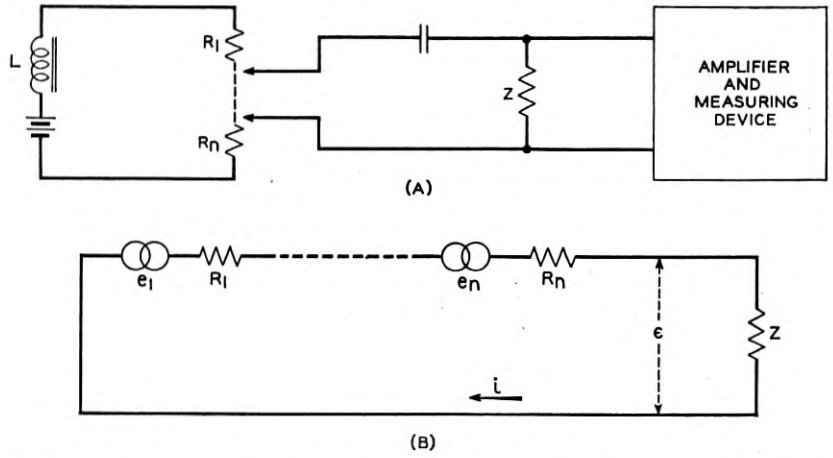


Fig. 13—(*A*) Circuit for measuring the contact noise of a series assemblage of resistance elements. (*B*) The equivalent a.-c. circuit of (*A*).

a.-c. circuit is shown in Fig. 13*B*. If the value of $\overline{\epsilon^2}$ impressed on the impedance Z of a measuring device is calculated, in a way similar to that outlined for the case of contacts in parallel, one gets

$$\overline{\epsilon^2}_{\text{series}} = \frac{[\overline{e_1^2} + \cdots \overline{e_n^2}]Z^2}{[R_1 + R_2 + \cdots R_n + Z]^2}. \tag{7}$$

Equations (6) and (7) are derived by considering the aggregate as made up of single contacts, but the same would be true if the aggregate were considered as made up of unit cells of arbitrary size, the values of e and R applying to the unit cells.

It is experimentally impracticable, if not impossible, to determine the noise and resistance of each contact in an aggregate; hence in the experimental test of the foregoing equations we have, in effect, divided the aggregate into unit cells. The aggregate of carbon granules was placed in four parallel grooves cut in a single phenol fibre block. Each groove was provided with five gold electrodes evenly spaced along the surface of the groove so that there were, effectively, four contiguous carbon cells in each groove. This permitted the measurement of the noise from each cell separately and also from the cells in various assemblages. Applying Eq. (6) or (7), as the case demanded, to the values of the noise from the single cells we calculated the expected noise of the assemblages and compared this with the measured values. Table I gives typical results for parallel assemblages and Table II typical results for series assemblages.

TABLE I

Cell	Cell Resistance Ohms	Noise in db	
		Measured	Calculated
a	3523	15.7	
b	4315	17.2	
c	4515	19.6	
d	3585	15.2	
$a + d$ parallel	1760	12.6	12.5
$b + c$ parallel	2215	15.7	15.5
$a + b + c + d$ parallel	980	11.2	10.9

TABLE II

Cell	Cell Resistance	Noise in db	
		Measured	Calculated
1	590	31.5	
2	720	34.0	
3	610	33.0	
4	510	30.0	
5	490	30.5	
6	545	31.0	
7	685	33.5	
8	427	26.5	
9	405	25.0	
10	226	19.0	
11	152	16.5	
Series 1 to 4	2510	38.5	38.2
5 to 8	2211	37.5	36.9
9 to 11	775	26.0	26.4
1 to 8	4720	41.0	40.4
5 to 11	2980	37.5	37.2

Considering the difficulty in holding a fixed granular configuration in a cell of loose contacts we feel that the experimental results justify the conclusion that the assumptions underlying the derivation of Eqs. (6) and (7) are essentially correct, which require that the phase of the noise voltage from each unit cell is entirely independent of that of any other cell in the aggregate. From this we conclude that the mechanism causing the noise is a small-scale effect capable of independently existing within a volume element much smaller than the size of the unit cell in any of the experiments we have performed on aggregates. In fact, as will be assumed later, we believe the noise mechanism to be located in a volume element smaller than that concerned with the properties of a contact between two particles.

If resistance elements are so chosen that each has the same resistance and noise, and these are placed in a circuit where the impedance Z is large compared to the resistance of the elements, then Eqs. (6) and (7) can be written respectively as follows:

$$\overline{V_c^2}_{\text{parallel}} = \frac{\overline{e^2}}{n}, \tag{6a}$$

and

$$\overline{V_c^2}_{\text{series}} = n\overline{e^2}. \tag{7a}$$

The resistance R of a parallel assemblage of like contact elements, each having the same resistance R_k, is obtained from $1/R = n/R_k$, or $n = R_k/R$.

Substituting this in Eq. (6a) we get

$$\overline{V_c^2}_{\text{parallel}} = \frac{R\overline{e^2}}{R_k}. \tag{6b}$$

For like contact elements in series we get $n = R/R_k$, hence Eq. (7a) becomes

$$\overline{V_c^2}_{\text{series}} = \frac{R\overline{e^2}}{R_k}. \tag{7b}$$

If now we have the further condition that the battery voltage is so adjusted for each new assemblage that $\overline{e^2}$ is always constant, then Eqs. (6b) and (7b) are equivalent and we have

$$\overline{V_c^2} = \text{Const. } R. \tag{8}$$

An equivalent relationship was derived and experimentally tested by Otto,[5] but it is clear from our derivation and measurements that it applies only to a change in the assemblage of like contacts, such as is

realized when the dimensions of a granular aggregate or conducting film are altered, and is not valid for cases where the resistance is altered by changing the contact compressions. In this latter case Eq. (2) applies.

Another interesting property of an assemblage is obtained if we express $\overline{e^2}$ by means of Eq. (1) in terms of the battery voltage V. Thus for the parallel assemblage, $\overline{e^2} = \text{Const. } V^\alpha$, and for the series assemblage, $\overline{e^2} = \text{Const.} \left(\frac{V}{n}\right)^\alpha$. Thus Eqs. (6a) and (7a) can be written, respectively, as follows:

$$\overline{V_c^2}_{\text{parallel}} = \frac{\text{Const. } V^\alpha}{n} \tag{6c}$$

and

$$\overline{V_c^2}_{\text{series}} = \frac{\text{Const. } V^\alpha}{n^{\alpha-1}}. \tag{7c}$$

If we now accept as an approximation $\alpha = 2$ then Eqs. (6c) and (7c) are equivalent, and we can say that for any assemblage of contacts, where the value of $\overline{e^2}$ for each contact element is equal to that of every other contact element in the assemblage, the contact noise of the assemblage is inversely proportional to the number of contact elements in the assemblage. This principle we have established experimentally by building "square" assemblages—$\sqrt{n}$ parallel paths with $\sqrt{n}$ elements in series in each path—and measuring the noise as a function of n. The "square" assemblage is particularly interesting for it allows a control of the noise of an assemblage without altering its overall resistance. This suggests a principle which may be followed in designing grid leaks and carbon transmitters with low contact noise characteristics.

Discussion

It seems to us that the most logical hypothesis consistent with the foregoing experimental data is, as before indicated, that the noise mechanism lies in a fluctuating contact or boundary resistance. Assuming this we are led to the following considerations concerning the nature of the noise mechanism.

Careful measurement has established that the conduction through a carbon contact, as near as can be observed, is entirely ohmic. We have shown that when a carbon contact through which direct current is flowing is cyclically compressed, as in the acoustic modulation of a carbon transmitter, the generated a.-c. power is proportional to the square of the d.-c. voltage. This leads to the conclusion that the

resistance modulation due to the cyclical compression is independent of the applied voltage. It is evident from Eq. (1) that the fluctuating resistance responsible for the noise cannot be equivalent to the resistance modulation introduced by a cyclical compression where the contacting granules move relatively as a whole, for the fluctuating resistance responsible for noise is somewhat voltage sensitive as indicated by the departure of α from the value 2. This means either that the conductance responsible for the noise is specifically non-ohmic or that the extent of the conduction wherein the noise mechanism lies is diminished as the applied d.-c. voltage is increased. Non-ohmic conductance is usually such that conductance increases with applied voltage, thereby demanding a value of α in Eq. (1) which is greater than 2; accordingly we are inclined to believe that applied voltage acts to diminish the area over which the noise mechanism operates.

If the noise mechanism were intimately associated with the total conductance of a contact one would expect the noise to be proportional to some simple integral power of the current in a contact, but this is denied by the observed value of β in Eq. (2).

These facts and deductions lead us to the hypothesis that there exist two mechanisms of conduction between particles in contact, a primary conduction which accounts for the major portion of the current, and a secondary conduction wherein a relatively small portion of the total current is transferred and in which the noise mechanism is found. Goucher [1] has given evidence that the primary conduction between contacting carbon particles is of the same nature as that in solid carbon, and since we have been unable to measure any noise in solid carbon we assume that the secondary conduction does not take place through the same region of the contact as the primary conduction.

Recent investigation of the elastic nature of carbon contacts [1] has led to the conclusion that the surface of each particle can be considered as covered with a layer of hemispherical hills of heights distributed according to the function $N_x = \text{Const. } x^n$, where N_x is defined as the quantity which when multiplied by dx gives the number of hills coming into coincidence with a plane as it moves from the position x to $x + dx$, and n is a constant whose experimentally determined value is about 0.6. The establishment of a contact consists in bringing into coincidence a number of these hills and enlarging the coincidence areas to the extent demanded by the displacement of the contacting elements after their initial coincidences. Let us accept this picture of a contact and inquire as to how it applies in the explanation of our empirical noise equation.

We assume that through each area of coincidence the primary con-

duction takes place and that secondary conduction, in which the noise mechanism lies, can take place between the surfaces which are not in primary contact and are not separated by more than a certain increment dx. Figure 14 is an attempt to picture a portion of the hypothetical plane of contact between two carbon granules.

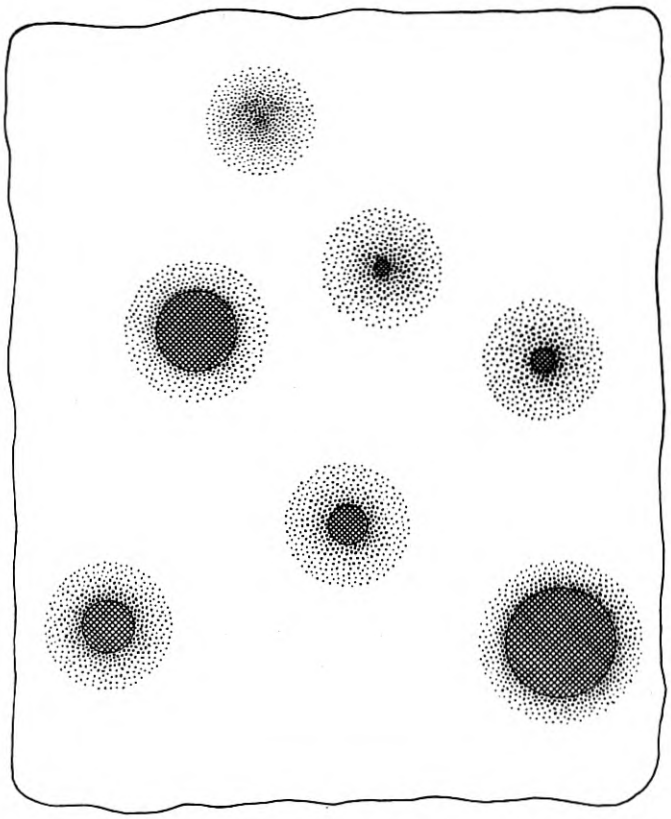


Fig. 14—A portion of the hypothetical plane of contact between two carbon granules. The cross hatched circular areas are the coincidence areas through which primary conduction takes place. The shaded areas, through which the secondary electrical conduction takes place, are regions where the granule surfaces are not separated by an interval larger than Δx. The area of each of these is independent of the coincidence area it surrounds.

The total number of hills in coincidence in a contact is given by

$$N_c = \int_0^D N_x \, dx = \text{Const.} \int_0^D x^n \, dx = \text{Const.} \, D^{n+1},$$

where D is the total displacement from the first coincidence. This number can also be expressed in terms of the contact resistance R by using Goucher's [1] derived equation: $1/R = \text{Const. } D^{n+3/2}$. Thus the above expression can be written

$$N_c = \text{Const. } R^{-(2n+2)/(2n+3)}. \tag{9}$$

The area of secondary electrical conduction surrounding each area of coincidence is precisely that area which would be added were the contact compressed by an increment of compression Δx. From the theory of Hertz [10] one can show that for smooth spherical hills in contact, $\Delta A/\Delta x$ is independent of the total hill compression and hence that the total area of secondary electrical conduction in a contact A_c is proportional to N_c, giving

$$A_c = \text{Const. } N_c = \text{Const. } R^{-(2n+2)/(2n+3)}. \tag{9a}$$

For purposes of analysis let us think of the secondary conduction area surrounding each primary area of contact as divided up into small elements of like nature, and further that the secondary conduction through each of these elements of area is independent of that in every other element. If such a contact is connected in a circuit as shown in Fig. 12*A* then the equivalent a.-c. circuit can be thought of as that shown in Fig. 15, where R is the mean resistance of the entire contact,

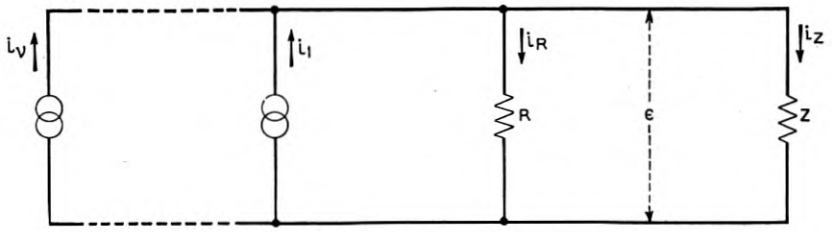


Fig. 15—The equivalent a.-c. circuit of a contact with ν elements of secondary conduction area, through each of which there is an independently fluctuating current, when such a contact is connected in a circuit as shown in Fig. 12*A*.

$i_1, \cdots i_2$ are the instantaneous deviations of the current from the mean current in each of the ν elements of secondary conduction area, and ϵ is the instantaneous value of the fluctuation voltage across the measuring device with impedance Z. Taking into account the random nature of the fluctuation currents through each element of secondary

[10] A. E. H. Love, "Mathematical Theory of Elasticity," 2nd ed., p. 192. This has been experimentally confirmed by J. P. Andrews, *Phys. Soc. Proc.* **43**, 1 (1931).

conduction area one can derive, in much the same manner as for Eq. (6), the relationship

$$\overline{\epsilon^2} = \nu \cdot \overline{i^2} \cdot (RZ)^2/(R + Z)^2, \tag{10}$$

where $\overline{i^2} = \overline{i_1^2} = \cdots \overline{i_\nu^2}$. Since ν is proportional to A_c we have by Eq. (9a) the relationship $\nu = \text{Const.}\ R^{-(2n+2)/(2n+3)}$. Substituting this into Eq. (10) we get

$$\overline{\epsilon^2} = \text{Const.} \cdot R^{(2n+4)/(2n+3)}\ Z^2/(R + Z)^2,$$

and if we make Z large compared with R this becomes

$$\overline{\epsilon^2} = \text{Const.}\ R^{(2n+4)/(2n+3)}. \tag{11}$$

Comparing this with Eq. (2), the experimentally derived relationship between noise and contact resistance, we get $n = 0.5$. Goucher [1] found by elastic measurements the value of $n = 0.6$.[11] In view of the fact that a slight change in the experimental value of the exponent β in Eq. (2) causes a rather large change in the value of n thereby determined from Eq. (11), we can say that the agreement between the results obtained from elastic measurements and noise measurements is surprisingly good, and that this agreement supports the hypothesis regarding the nature of a contact.[12]

In discussing the hypothesis that there exists a region of secondary conduction which is responsible for the noise, we have not made any assumption as to the nature of the secondary conduction. Several possibilities, however, have occurred to us, one of which it seems desirable to mention at this time.

If one assumes that the thermo-mechanical vibrations of a solid extend to the outside surface, then it is possible that the wave crests may be able to make periodic electrical contact across the secondary conduction area assumed in our hypothesis. This would permit a pulsating current to flow, the frequency of which, it is supposed, is determined by the frequency of the oscillator. For oscillators of audible frequencies the law of energy equipartition applies and each oscillator will have the usual $1/2\ kT$ of energy per degree of freedom. The energy of an elastic oscillator is also proportional to $(B \cdot F)^2$, where

[11] Goucher found a discrepancy between the measured resistance-displacement, resistance-force, and force-displacement relationships. But for reasons stated in his paper we are inclined to accept the distribution function found from the force-displacement measurements.

[12] If one assumes the existence of a film in the contact, which some older theories of microphonic action do, and that the number of independent elements of area through which current is conducted is proportional to the area of this film then one is led to the very unsatisfactory conclusion that $n = -3.5$.

B is the amplitude and F the frequency. From these two expressions of the energy we get $B \sim F^{-1} T^{1/2}$. The area of secondary electrical conduction surrounding each hill in coincidence is proportional to B, and since this area determines the number of elements of area through which secondary conduction takes place, as assumed in the derivation of Eq. (11), we arrive at the conclusion that

$$\overline{\epsilon^2} = \text{Const. } F^{-1} T^{1/2}. \tag{12}$$

While the hypothesis leading to Eq. (12) is only intended as a suggestion it does explain the inverse frequency relationship, and the temperature relationship is not an impossible one judging from the past unsatisfactory measurements. It may be possible, also, to explain the departure of α in Eq. (1) from the value 2, for one would expect electrostatic forces to distort the contacting surfaces so that the secondary conduction area would become smaller as the voltage increases. A satisfactory experiment on the effect of temperature on noise will do much to establish or disprove the tenability of this hypothesis.

Brillouin [13] has recently derived an expression for the noise in a conductor carrying a current by using the statistical method to deduce the most probable distribution of the electrons in such a system when it is in equilibrium. This method of calculation gives a noise energy, in addition to thermal noise, which is proportional to the square of the current and inversely proportional to the volume of the conducting material. We have made a calculation of the relative magnitudes of the two terms in Brillouin's equation which correspond to our experimental conditions. Assuming reasonable dimensions for a carbon contact and a current density as high as any we used it turns out that the magnitude of the term for contact noise is far below that for thermal noise. Furthermore it seems to us that Brillouin's mechanism would require a flat frequency distribution of noise rather than the distribution which we have observed. For these reasons we do not believe that the noise which we have studied is produced by the mechanism postulated by Brillouin.

In conclusion we wish to acknowledge our indebtedness to Dr. J. B. Johnson and Dr. F. S. Goucher for the helpful criticism they have given us during the course of this work.

[13] L. Brillouin, *Helv. Phys. Acta, Supt.* 2, **7**, 47 (1934). This theory is intended to explain the "Fluctuations de résistance dans un conducteur métallique de faible volume," reported by M. J. Bernamont, Comptes Rendus **198**, 1755 (1934); ibid. **198**, 2144 (1934).

Contemporary Advances in Physics, XXX—The Theory of Magnetism *

By KARL K. DARROW

The topic of this article is the explanation of magnetism as ordinarily observed—to wit, the magnetization of pieces of matter of ordinary dimensions—by ascribing magnetic moment to the individual molecules, atoms, and electrons of which matter is composed. For *paramagnetic* bodies it is postulated that the individual atoms are magnets of which the orientation, but not the strength, is altered in the presence of a magnetic field; the theory is so successful as to make it possible to calculate, from magnetization-curves, values for the magnetic moments of these atoms which agree admirably with those deduced from spectroscopic theory and from experiments of other types. For *ferromagnetic* bodies the same postulate is made, but it is necessary in addition to recognize the existence of huge interatomic forces of which very little is known, so that a large proportion of the science of ferromagnetism still lies beyond the scope of atomic theory. For *diamagnetic* bodies the phenomena are interpreted in a simple and effective manner, as an immediate corollary of the well-known structure of the atom.

MAGNETISM is a quality which we attribute to the atom. We affirm that iron, nickel, gadolinium, gaseous oxygen, and in fact all substances, are magnetic because there is magnetism in their atoms. Indeed we go even deeper, and affirm that the individual electrons and the nuclei within the atoms are magnetic. Nevertheless, the atomic theory of magnetism is a really valuable theory. Perhaps that "nevertheless" sounds out of place; but I assure you that without it there would be a trace of paradox in the statement, which perhaps our grandfathers would have been quicker at discerning than are we. Let me explain my meaning by referring to the atomic theory, or as it is usually called the kinetic theory, of gases. Those who designed this theory succeeded in explaining the pressure, the temperature, and the viscosity of gases, without attributing a single one of those qualities to the atoms. To the atoms they assigned the properties of momentum and velocity and kinetic energy; those other qualities which I just named were then interpreted in terms of these,—they were interpreted as what we call *statistical* properties of the great multitude of atoms which constitutes a gas. This was a real explanation of pressure and viscosity and temperature, in the fullest sense of the word "explanation"—or anyhow, in the fullest sense of that word which is customary in physics. But along with these properties of pressure and viscosity

* Expanded from a lecture delivered on January 14, 1936, at the School of Engineering of Yale University, and still bearing obvious traces of its original form. In preparing it I received invaluable aid from Dr. R. M. Bozorth.

and temperature, a gas also possesses weight. The builders of the kinetic theory simply said that the weight is a property of the individual atoms, and that the weight of the gas is the sum of the weights of its atoms. Now evidently this was not an explanation of weight at all. Indeed, by assigning weight to the individual atom, the builders of the theory had foregone all attempts at an explanation. A property which you assign to the atom is a property which you refuse to try to explain in terms of the atom—or so at least it always seemed to our forefathers. To assign a quality to an atom used to be taken as a confession of incompetence to explain that quality. I can of course make this clear by proceeding to absurd extremes. If I say that an orange consists of soft yellow juicy atoms, or that a marshmallow is made of sweet white sticky atoms, or that a piece of iron is made of hard black shiny conductive atoms, you recognize at once that those are not serious atomic theories: they are just futile and somewhat ridiculous statements. If I claim to explain the weight of a piece of iron by saying that it is the sum of the weights of the atoms, I am making a claim which unfortunately may not sound ridiculous, but is really just as futile—unless it acquires value by being linked with some other assertion. But when I say that the magnetism of a piece of iron is due to the magnetism of its atoms and its electrons, the statement is by no means a futile one; it is significant and important. For this there are two main reasons or rather groups of reasons, which I will indicate by the words *orientation* and *atomic structure*. (In addition there are remarkable experiments on jets of atoms whereby their magnetic moments are measured directly, but these I reserve for another occasion.)

First a few words about atomic structure. It is a fact of experience —the experience of one hundred and fifteen years—that a current running around a loop of wire is the equivalent of a magnet. If now somebody asserts first that a piece of iron is magnetic because its atoms are magnets, and then goes right ahead and asserts that the atoms are magnets because they have perpetual currents running around inside them—well, the combination of these two statements is not necessarily futile or trivial. At the very least, it is a sensible attempt to reduce the two kinds of magnetism apparently existing in the world to a single kind, that which is due to moving electricity. This was Ampere's idea a hundred years ago. Now if in addition there is independent evidence that the atom comprises mobile electrical particles, then this idea of Ampere's becomes the assertion that those particles inside the atom are actually revolving. It is well known that modern physics is full of such evidence, of evidence that atoms contain very mobile electrons; and some of my readers may recall that thirty years or so ago there

were an atom-model with stationary electrons and an atom-model with revolving electrons, which were in competition with each other, and the latter of which has by now driven the former utterly out of the field. Remember now, that the atom-model with the revolving electrons triumphed over the other one not primarily because of its magnetic quality, but because of the theory of spectra which Bohr and others were able to derive from it. Revolving electrons in atoms were first of all proved to be responsible for spectra, and then it was noticed that they are capable of causing magnetism. Therefore when the physicist says that magnetism is a quality of atoms, he is not making a confession of incompetence, but an inference from a highly-developed and successful theory of spectra; and this makes all the difference in the world to the value of the statement. Indeed the situation is even better than I have intimated; for there are dozens of cases in which first an atom-model or a molecule-model has been constructed expressly to explain the spectrum of the substance in question—then, the magnetic moment of its system of revolving electrons has been computed—then, the magnetic moment of the atom or the molecule has been measured—and the two have agreed! This is really an understatement, which needs to be broadened so as to include the cases in which the spin of the electron plays a part; but I pass them over, intending to defer the broadening to the latter part of the article, which is to be devoted to these matters of atomic structure. For the moment, let me make just one more allusion to them, a very important one. Electrons revolving in orbits around a nucleus obviously possess angular momentum. Therefore, if an atom has a magnetic moment due to revolving electrons, it has an angular momentum also. This again is an understatement, for it contains a restriction which can be removed in view of the broadening which is later to be made. It appears to be a general rule that *in the atom, magnetic moment and angular momentum always go together.* A magnetic atom is a gyroscope—necessarily and automatically. This is a fundamental principle, and from it flow some strange and striking consequences. Everyone who has worked or played with the classic gyroscope of our laboratories knows that it has quaint and tricky idiosyncrasies. Well! the atom has them too; but it has others in addition. Angular momentum, on the atomic scale, is subject to peculiar laws of quantum mechanics; and the atomic magnet-gyroscope behaves in extraordinary ways, of which our laboratory gyroscopes give not the faintest intimation.

To summarize my introduction then: the first step in the theory of magnetism consists in referring it to the individual atom. This sounds like a confession of defeat, but it is nothing of the sort; it is a claim of

victory. Our theory of spectra requires that atoms, or some of them at any rate, should be magnets, and they *are* magnets. Moreover it fixes the magnetic moments which certain atoms ought to have, and so far as our experiments go, the atoms *do* have these moments. Moreover it imposes angular momentum on these atoms, and fantastic as the consequences are, experience bears them out. Logically, then, I should begin the main part of my talk by showing how the magnetic moments and the angular momenta of atoms and of molecules are calculated from their spectra by atomic theory. This, however, would by itself require several lectures, and very difficult ones at that.* I must therefore simply ask you to believe that the magnetic moments of atoms are inferences from fundamental theory, not mere *ad hoc* assumptions; and now I will explain what I had in mind when I wrote down the word *orientation* to designate one of the topics of this article.

It is one of the best-known facts of physics that the magnetization of a substance is not fixed and constant, but increases with the strength of the magnetic field which is acting on the substance. By the way, before going any further I must definitely exclude the so-called "diamagnetic" substances. That exclusion being made, we do *not* assume that the magnetic moment of the individual atom increases similarly with the field strength. People did not make that assumption, even in the days before the fundamental theory was developed. Had they done so, it would have been just as silly as saying that a marshmallow is made of soft white sticky atoms, and calling *that* an atomic theory. They supposed, and we suppose, that the moments of the individual atoms remain practically the same whatever the field strength; what changes is the average *inclination* of these moments to the field. The atomic moments are vector quantities pointing in various directions, different from one atom to the next. The magnetization of the substance as a whole is the resultant of all these myriads of tiny vectors pointing in their various directions. If they all pointed the same way the substance would be completely and perfectly magnetized, with a moment equal to the total number of the atoms multiplied into the moment of any one. This state of *saturation* is not, however, to be attained, not even to be approached without a rare and felicitous concourse of a favorable substance, a very low temperature and a very strong field. Much easier of attainment is the opposite extreme, when the vectors are pointing all ways at random and the magnetization is zero. This happens with nearly all substances when there is no field applied, and it seems quite natural. But when even the smallest field

* This subject was partially treated in "Contemporary Advances in Physics, XXIX . . . ," April 1935 *Bell Sys. Tech. Jour.*

strength is applied to such a substance, you might expect all the little magnets to turn right around and point straight up the field-direction, achieving saturation in an instant. Well, it is certain that saturation is not achieved; but still there is some degree of magnetization, as though the little magnets all started to turn around and were stopped before they got very far. What is it that might stop them? If you look at the books of twenty or twenty-five years, you will find an answer: they are stopped by the collisions which these atoms make with one another. This is the classical idea, which is generally thought to be well verified by experiment. But let us look into the matter a little more closely.

For simplicity let us imagine a gas—preferably, unit volume of the gas—composed of N identical atoms, each with a magnetic moment μ, and exposed to an applied field H. Visualize some particular atom, of which the career is an endless alternation between free flights and sudden impacts. All the time the magnetic moment of the atom, the little vector of which I have been speaking, is subject to a torque arising from the field. The classical idea is, that throughout every free flight that torque is steadily bringing the vector more and more nearly into alignment with the field, but usually not having time enough to succeed, because at every collision the vector is suddenly and violently redirected in a perfectly arbitrary way. Gradual approach to alignment during the free flights, violent dis-alignment at the collisions, and the magnetization of the substance indicating how far the alignment progresses, on the average, before the dis-alignment stops it—this is the classical picture. It all seems beautifully obvious, and yet is it now believed to be entirely false!

The trouble lies in the fact that the atom is a gyroscope. You recall that it is one of the oddities of the gyroscope that when you apply a torque to it, it starts off at right angles to the direction in which you expect it to go. Now here is our atom just leaving the scene of a collision with its magnetic moment making, say, an angle ϕ with the field-direction. As it flies away the torque is steadily trying to reduce the value of ϕ, but instead of obeying, the atom just keeps on blandly precessing about the field-direction, the value of ϕ remaining obstinately the same. The unbreakable link between magnetic moment and angular momentum has neatly killed the supposition that the field magnetizes the gas because it aligns the atoms, or partially aligns them, during their intervals of unimpeded flight. The free flights are just the periods when nothing whatever happens in the way of alignment. Much labor has been expended in the hope of finding some way out of this impasse, but none has been revealed except that of supposing that

whatever is the mechanism whereby the field aligns the atoms, it is a mechanism which operates during the impacts and not between them. Partial alignment *at* the collisions, no change in the situation during the free flights—this amounts pretty nearly to standing the classical theory on its head!

Nevertheless the mathematics of the classical theory remains entirely unchanged. This is because the mathematics merely expresses the assumption that the field has managed to find *some* way of partially aligning the atoms, and does not concern itself in the least with what that way may be. This sounds rather vague, so let me remind you just what the assumption is. Suppose to begin with that the vectors of the atoms are capable of only two orientations in the applied field: one parallel, the other anti-parallel to the field-direction. To transfer an atom from the one orientation to the other, we must do work against the torque of the field (or receive work from the torque of the field) amounting to $2\mu H$. We have, therefore, two classes of atoms, differing in energy by $2\mu H$. Let N_1 and N_2 stand for the numbers in these classes at some particular instant. Now the classical theory, as I have been calling it, is strictly no more than the assumption that the ratio of N_1 and N_2 is given by Boltzmann's theorem:

$$N_1/N_2 = \exp(-2\mu H/kT) \tag{1}$$

and the essence of this assumption, I take it, is that the atoms are able to change their orientation so as to pass from either class to the other, and that they employ this facility of free passage to get themselves into thermodynamic equilibrium at the temperature T of the gas. This has been the assumption ever since Langevin founded the theory, and it still is the assumption, even though we may no longer enjoy that pretty picture of the mechanism of the change of orientation which once we accepted, and have no other to replace it.

I can readily write down the complete theory of this case. We introduce the two additional equations,

$$N_1 + N_2 = N, \qquad I = (N_1 - N_2)\mu, \tag{2, 3}$$

of which the first says that all the atoms are in either the parallel or the anti-parallel class, and the second that the magnetization I of the unit volume of gas is the resultant of the vectors of all its atoms. Now we eliminate N_1 and N_2 between the three equations, and swiftly arrive at the result:

$$I = N\mu \tanh(\mu H/kT), \tag{4}$$

which is the equation of a curve starting obliquely off to the right from

the origin, and bending over to approach an asymptote which is a horizontal line of the ordinate $N\mu$.

At this point a strictly classical physicist would certainly grin or sneer, because he would say to himself: "The speaker started out by assuming for simplicity that the atoms can point in only two directions, and now he has gone on to his conclusions without remembering the obvious fact that an atom may point in any direction whatever!" Well, of course Langevin did make allowance for that supposedly obvious fact; it complicates the affair to some extent, but not seriously, and leads to a very similar curve for I versus $\mu H/kT$. Quantum mechanics, however, flatly denies that it is a fact. I mentioned above that the atomic gyroscope has some paradoxical properties of its own, in addition to those which it shares with the laboratory gyroscope. Here is one of them. *The atomic magnet is supposed to be able to set itself*, not at any angle whatever with respect to the applied magnetic field, but *only at one or another of a small number of definite discrete angles*. This is because of its angular momentum: it is primarily the angular momentum which is constrained to this very singular behavior, and which the magnetic moment is automatically obliged to follow because they are so closely linked together. If I am asked why the angular momentum should behave like this, I can only reply that according to what I am told, if one is sufficiently penetrated with the spirit of quantum mechanics it seems self-evident, and if one is not sufficiently penetrated with that spirit there is nothing which can be done to help. Notice anyhow that it is compatible with the statement that when the atom is freely flying along, the field just keeps it precessing about the field-direction, instead of gradually aligning it; and there is ground for being thankful that this derivation, and certain others, are somewhat simplified by it. It may be asked, how many different inclinations are permitted to the atom? This depends upon the angular momentum of the atom, and we can tell it from the spectrum. There are certain elements and certain compounds for which the case is just as simple as I have described it; just two permitted orientations, the parallel and the anti-parallel, and no more. There are others for which the permitted inclinations are three in number, others for which there are four, five, and other integers up to fifteen or twenty. All these yield curves of I versus $\mu H/kT$ having the same general traits, but differing in the rate at which they approach the asymptote. I will refer to all such curves as "Langevin curves," although the only one which Langevin proposed was the classical curve corresponding to the case in which all orientations are permitted (or, as we may say, there are infinitely many permitted orientations).

You may now be expecting me to say that there are many gases, and possibly other substances as well, for which experimental curves have been obtained that are comparable with these. I am obliged to disappoint you. You can readily see that in order to get over onto the "curvy" part of these curves, one must work in experimental conditions in which the argument $\mu H/kT$ is greater than, or anyhow not very much less than, unity. One thinks, of course, of using the highest accessible field strengths H so as to enhance the numerator of that fraction. This, however, is not sufficient, for it turns out that μ (the magnetic moment of an atom or a molecule) is so very small that one is obliged to diminish the denominator also by going to the lowest attainable temperatures. All the experimental curves of this character have been obtained at temperatures lower than 15° absolute, some at temperatures between 1° and 2° absolute. This excludes all the gases. Moreover, it has been necessary thus far to choose the atoms with the largest magnetic moments, and these turn out to be, quaintly and inconveniently enough, the atoms of the rare-earth elements. Probably the best of the experimental curves (Fig. 1) relates to a substance which most people never have heard of, in this or any other connection: it is gadolinium sulphate. There are about a score of such curves ob-

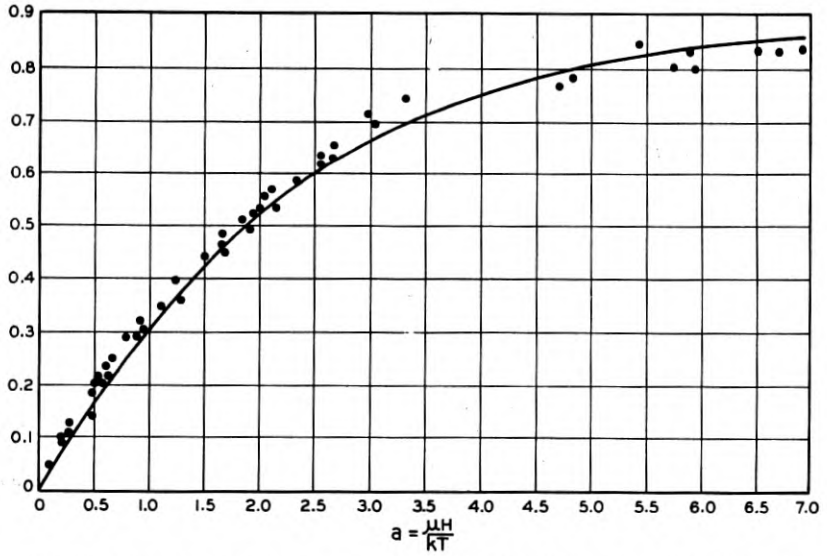


Fig. 1—Magnetization of a paramagnetic salt $[Gd_2(SO_4)_3 + 8H_2O]$ as function of the parameter a; the ordinate is I referred to its saturation-value (deduced by extrapolation) as unity. Data from Onnes and Woltjer. The curve is the "classical" Langevin curve (number of permitted orientations, $n = \infty$) which is hardly distinguishable from the quantum-theory curve for this particular case ($n = 15$).

tained with minerals and glasses containing these elements, and most of them agree well with one or another of the theoretical curves; in which connection there is an interesting detail, which I will bring up at the end of the article. One would scarcely expect a theory worked out for gases to apply so well to solids, and as a matter of fact it is a peculiarity of the rare-earth atoms that even when incorporated in a compound or a solid they behave in several ways more like the atoms of a gas, than happens with any other elements.

Pray do not think, however, that all this time I have been talking about a theory which has no application excepting to the rarest of all elements under the rarest of all temperatures. Its applications are a good deal wider than that. True it is that with gases universally, and with other substances ordinarily, we cannot get data along the curvy parts of the curves; but we can make measurements along the sensibly-linear parts near the origin. This amounts to saying that we can determine the slope of the curve at the origin. Now of course it sounds ridiculous to speak of confirming a theoretical curve by measuring its tangent at one point. In this case, however, it is not altogether ridiculous. Usually the experiments are made by varying H and measuring I while the temperature is kept constant. Suppose this is done for several different temperatures, and suppose the results are plotted by using H instead of $\mu H/kT$ for the abscissa. Then the theory supplies us with different curves for the different temperatures, all having the same general aspect, but different slopes at the origin. I will denote these slopes for the time being by $\tan \theta_0$. The theory, then, requires that $\tan \theta_0$ should be proportional to $1/T$; and for gases, this is found to be the case. Of course this is not such good evidence for the theory as would be a complete following-up of the curve nearly all the way to the asymptote; but it is pretty good by itself, and for further evidence we can invoke those experimental curves for gadolinium sulphate and other solids of which I just spoke.

If now we let ourselves be convinced by this evidence, a valuable conclusion follows. From the slopes of these curves at the origin, the value of μ can be deduced. Let us go back to the curve of I versus $\mu H/kT$ or a, which is the epitome of all the rest. We write:

$$dI/da = N\mu(1 - \tanh^2 a), \tag{5}$$

$$(dI/dH)_{T=\text{const.}} = (dI/da)(da/dH) = (1 - \tanh^2 a)N\mu \cdot (\mu/kT). \tag{6}$$

Since measurements are actually made at a fixed temperature and refer to the slope of the curve near zero field strength, we evaluate this derivative for $a = 0$, and we get:

$$\tan \theta_0 = (dI/dH)_{H=0} = N\mu^2/kT, \tag{7}$$

and thus from the measurement of θ_0 at any temperature we derive the magnetic moment of the individual atom or molecule of the gas. This formula ought to give the right order of magnitude for μ in any case. Whether or not it will give exactly the right value, will depend on the validity of one of the assumptions, which I now recall. This particular formula is for the case in which the atoms have only two permitted orientations in the field, the parallel one and the anti-parallel one. Had we supposed that every inclination is a permissible one, we should have arrived at $(1/3)N\mu^2/kT$. Had we supposed a number of permitted inclinations greater than two and less than infinity, we should have arrived at some intermediate value. So, I now write as the general formula,

$$\text{volume-susceptibility } \chi = bN\mu^2/kT, \qquad b = 1 \text{ to } 1/3, \qquad (8)$$

having placed on the left the name and the symbol by which is generally known what I have been denoting by *tan* θ_0, and on the right a factor b of which the value will depend on the number—I will call it n—of permitted orientations, but will fortunately never be outside of the narrow range between unity and 0.33.

Thus a rough estimate of an atomic moment may be made without knowing the number of the permitted orientations. Very many such estimates have been made, and they always give values of μ quite compatible with what we know in general about the structures of the atoms. If we want to make an estimate truly accurate enough to serve as a stringent test of theory, then we must take from the spectrum of the atom, not only the spectroscopic value of magnetic moment with which we are going to make the comparison, but also the angular momentum of the atom which is what determines the number of orientations. This causes us no extra trouble, for if we understand the spectrum well enough to get the one we also understand it well enough to get the other. Now when we look into the literature to see how many such comparisons have been made, we suffer again a disappointment. It turns out that the noble gases and most other convenient gases exhibit the magnetic moment *zero*. This is of course no fortuitous bit of ill luck; it is the same thing, to wit a certain stable interlocking of the various electronic orbits and rotations in the atom, which leads on the one hand to a zero magnetic moment and on the other hand to a relative smallness of the forces which make for chemical combination and for condensation. Anyhow it is an inconvenience; but luckily there are two convenient gases, oxygen and nitric oxide—O_2 and NO—which do have magnetic moments different from zero; and the test of the theory is in these cases most satisfactory. The agreements be-

tween the magnetic moments calculated from magnetic data after this fashion, and those derived from the spectra, are accurate within an experimental uncertainty of a few promille. I think that these are among the most impressive results in the whole structure of modern physics. Then in addition the rare-earth elements help us out again, owing to the peculiarity which their atoms have of behaving, even when they are incorporated into solid compounds, as though they were the atoms of a gas. They have supplied us with a number of beautiful agreements of this same character.

Now as a transition to the next part of this paper, I must acquaint you with another fact which belongs to this last part. I have more or less been allowing you to suppose that with solids as with gases, the susceptibility is generally proportional to $1/T$. Actually it is much more common, among solids, to find a law of the type,

$$\chi = \text{const.}/(T - \theta), \tag{9}$$

where θ stands for a constant differing from one substance to another. This constant is evidently of the dimensions of temperature; it is a sort of "critical" temperature, known as the *paramagnetic Curie point;* the formula usually holds for a broad range of values of T above and not too close to θ. (There are plenty of cases where even this formula will not fit, but we will not concern ourselves with them.) You see that this might be taken as meaning, that for temperatures greater than θ the substance is more strongly magnetized by any particular field strength than, by our previous theory, we should expect it to be. It might even be taken as suggesting, that in addition to the applied field which we produce ourselves by a horseshoe magnet or something of the kind, there is an extra field arising within the substance itself, which helps along with the magnetization. Now this is just the suggestion which physicists have accepted. Of course it is necessary to make some specific assumption about this extra or internal field, in order to arrive at the empirical law which I just wrote down. The required assumption turns out to be simple and gratifying. It is necessary and sufficient to assume that inside the magnetized substance, there arises an extra field which is proportional to the magnetization I itself. Hitherto we have been supposing that the torque acting upon an atomic magnet is directly and entirely due to the applied field H, and we have been led to the law that χ varies inversely as T. Now we are going to suppose that the torque is due to a field $(H + AI)$; and this will lead us, by way of the equation

$$I = N\mu \tanh (H + AI)/kT \tag{10}$$

to the law that χ varies inversely as $(T - \theta)$. The constant A is one which we adjust so as to get the empirical value of the constant θ. It is pleasant to be able to say that this constant A—that is to say, the hypothetical extra field—does not meddle at all with the multiplying factor: the theory allows us to write:

$$\chi = \frac{bN\mu^2}{k(T-\theta)} \quad \begin{matrix} \theta \text{ depending on } A \\ b \text{ depending on } n \end{matrix} \tag{11}$$

and (if it is the right theory, of course!) we can go on estimating atomic moments for substances of this category just as well as we can for the substances for which θ is zero. Most published values of μ correspond to such cases.

I am going to say very little about the extra field, or "Weiss field" as it is often called, because it is still one of the mysteries of physics. One realizes readily, of course, that if all the little atomic magnets turn themselves partially or totally into alignment, each one of them experiences a magnetic torque which is due to all the rest. It may be shown that this is proportional to the magnetization I, which looks very promising indeed; but alas, when it is calculated its magnitude turns out to be thousands of times too small. People used to say that AI must be a field of non-magnetic origin, which is just another way of saying the same thing. At present it is commonly believed that the force in question is what is called an "exchange" force, that is to say, an electrostatic force among electrons, of which the *modus operandi* can be discovered only by quantum mechanics. I am told that this quantum-mechanical theory has not yet been persuaded to deliver a really satisfactory result; but probably we shall be obliged to accept it in default of any other.

Now I call your attention to the fact that if the temperature should be made equal to or lower than θ, this last equation would predict something very wild and strange: an infinite, or a negative, susceptibility. This is a curious situation, and there are several cases in which we can appeal to experiment to resolve it. Take the elementary metal *nickel*, for example; if one measures the susceptibility over the range between 400° and 900° C. one gets a gorgeous curve of just this character, for which the value of θ is around 370°; now if one investigates nickel at temperatures below 370°, say around room-temperature, one learns that it is *ferromagnetic*. The same holds true for iron, for cobalt, for a diversity of alloys, except that θ varies from one case to another.*

* There are however cases in which the substance does not display the distinguishing marks of ferromagnetism (notably remanence) when $I < \theta$; and incidentally there are cases in which θ is negative; all of these are knotty problems for theory.

I will just refer to one additional case, because it is of very recent discovery and relates to that rare element which is so helpful in magnetics and seems to be so useless for anything else: I mean *gadolinium*. Metallic gadolinium has a value of θ amounting to about 300° absolute. Well, last spring Trombe at Strasbourg investigated this metal at low temperatures and found that it, too, is ferromagnetic, even more so than iron itself. Incidentally most of the rare-earth elements have not yet been prepared in pure metallic form, and it looks as though we might almost count on turning up some more cases of this kind. All this brings me to the question of *ferromagnetism*.

I do not suppose that any of my readers thinks that it is ferromagnetism of which I have thus far been speaking, but for the sake of completeness I will give the name: up to this point we have been considering *paramagnetic* bodies, and explaining their behavior by the orientations of atoms in fields. Now we turn to the properties of iron, cobalt, nickel, various alloys and compounds of these, various alloys containing manganese, and gadolinium: the *ferromagnetic* substances.

The most confusing thing about ferromagnetism—at least if my own experience as a student is any guide—the most confusing thing is, that the I-vs-H curve of a ferromagnetic substance reminds one of the sort of thing that the Langevin theory is meant to explain, and yet it is not that sort of thing at all. One looks at the Langevin curve with its approach to saturation, and then one thinks of the curve for iron with *its* approach to saturation, and one cannot help but think that the two must correspond to each other except for minor and trivial details. Well, they do *not*. They differ not alone in trivial details, but in every possible way, excepting the solitary common feature of the approach to a horizontal asymptote.

It is really impossible to put this statement too strongly. The Langevin curve and the iron curve differ in shape, as any sketch (cf. Figs. 1 and 2) will show. They differ utterly in scale. If I were to start to put a Langevin curve on the same plot where an iron curve appears with suitable detail, not only would it be sensibly linear for thousands and thousands of miles, but it would not even rise appreciably off the axis for hundreds of miles. Conversely if I had tried to put the curve for a ferromagnetic body upon the same graph as the Langevin curve, the former would have consisted only of the axis of ordinates plus the horizontal asymptote. Finally, the temperature relations are all wrong. I told you that in the Langevin curve the slope near the origin varies inversely as temperature, and I left you to infer that the ordinate at saturation is independent of temperature. In the curve for iron, the slope near the origin goes up with the temperature, and the

ordinate at saturation goes down as the temperature goes up. And yet, we interpret ferromagnetism by what is essentially an atomic theory: that is to say, we suppose that any piece of iron is an aggregate of little magnets each having a constant magnetic moment (so long as the temperature is kept constant) and that magnetization of iron consists in aligning these magnets.

I think it instructive to refer to these little magnets by the name of "atom," with some distinctive prefix; so, for a few minutes, I will call them "super-atoms," though this is not the customary name. When a piece of iron is unmagnetized or demagnetized, the super-atoms are

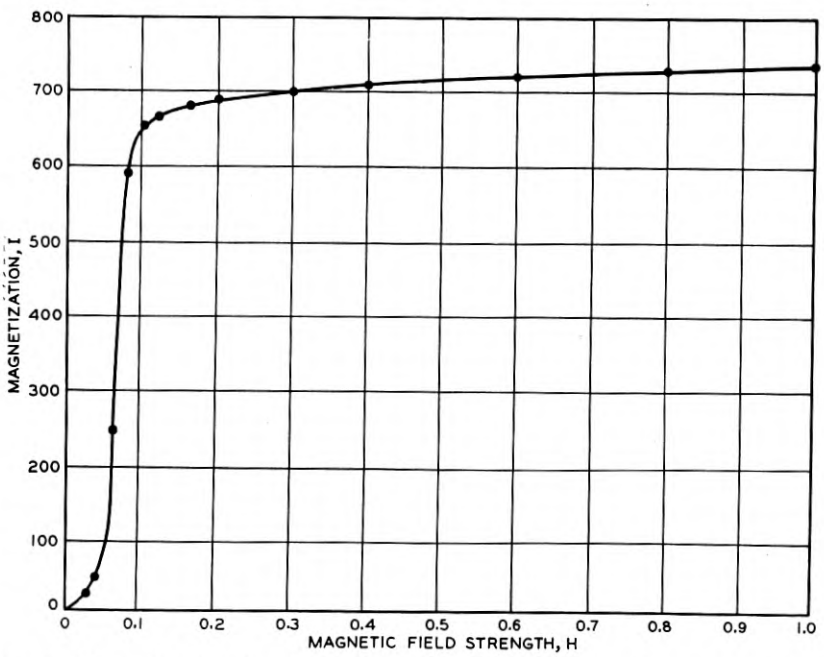


Fig. 2—Magnetization of a ferromagnetic material (81 permalloy, annealed two minutes at 1000° C.); the ordinate is *I*. Data by P. P. Cioffi.

pointing in all directions at random, just like the individual atoms of a paramagnetic gas which is unmagnetized. When a magnetic field is applied to the unmagnetized iron, the super-atoms get more or less aligned with one another. If the field is strong enough they are perfectly aligned, and there exists what is usually called "saturation" of the iron. Now it is worse than useless to remember about Boltzmann's theorem, or impacts, or free flights between impacts, for all those concepts have no relevance. We have to look at the phenomena, and see what they require.

We see at once that the super-atoms must be very easy to align, because saturation comes so quickly, with so relatively small a field strength. We learn also that when they are aligned, they are not exposed to the incessant urge to utter dis-alignment which afflicts the atoms of a paramagnetic substance, for iron continues to be magnetized when the field is withdrawn; not fully magnetized, as a rule, but considerably so. Heretofore I have been talking of substances, in which the atoms have a natural state of perfect dis-alignment or random orientation; a moderate field can derange it only a little, and the atoms return to it instantly and invincibly as soon as the field is cancelled. Now I am talking of substances in which the super-atoms seem to have no single natural state at all; a moderate field aligns them with ease, and when it is removed they like to linger in their alignment. The phenomena become clearer when we experiment not with ordinary iron, which is a chaotic mass of tiny crystals, but with a single large crystal. It turns out then that the super-atoms have a mighty preference for pointing along the cubic axes as distinguished from all the other directions; but as between these three cubic axes, and as between the two opposite senses along each of the three, they seem to be well satisfied with any. Suppose for definiteness that I have a cubic crystal of iron with one of its axes vertical, another in the meridian and the third, of course, pointing east and west. Then if the crystal is unmagnetized, one sixth of the super-atoms may be pointing east and one sixth west, one sixth pointing north and one sixth south, one sixth pointing up and one sixth down. (I do not say that this is necessarily the case, but it may be.) Now if I apply to the crystal a moderate magnetic field pointing north, the one sixth of the super-atoms which were already pointing north will not be affected, but all the other five sixths will flop right over and imitate them. It is amazing how small a field will suffice to do this: 100 cersteds for a good single crystal, whereas 100,000 oersteds, as I suggested, are not enough to bring the ordinary paramagnetic substance at room-temperature anywhere near to saturation. If next I cancel the field, the five sixths of the super-atoms which came over to the northward orientation will not be irrestibly urged to hasten back to their previous habit: indeed if I manage to avoid mechanical shocks and jarrings, most of them may linger indefinitely, still pointing in the direction to which the vanished field once tempted them. Some readers may notice an odd resemblance between this and the earlier case, in that the super-atoms have a finite number of discrete orientations, just as the atoms do. This resemblance is, however, so superficial and (probably) misleading, that I might not even mention it if I could be sure that it had not been

observed. To state two points of difference among many: the "permitted" directions for the super-atoms depend upon the crystal structure, those for the atoms depend upon the field-direction and the angular momentum of the atom; and if one applies a field to a single crystal in any direction oblique to all of the cubic axes, the super-atoms will consent to point in that direction, provided the field strength is rather high.

Now I must explain what these super-atoms are, since our understanding of them is one of the most satisfactory features in our, as a whole very imperfect, theory of ferromagnetism. They are groups—commonly called *domains*—of adjacent individual atoms; the member-atoms of each domain are behaving like the atoms of a paramagnetic solid. A diagram of a ferromagnetic solid might be drawn as an assemblage of large arrows, each representing the magnetic moment of a single domain; then, around and beside each of these large arrows might be drawn a lot of small arrows representing the magnetic moments of the individual atoms constituting the group; the big arrow would be the resultant of all the little ones. It would not be practicable to do this accurately, for there would have to be millions, or millions of millions, of little arrows to each of the big ones; but even a few suffice to show the idea. It may, however, be recalled that I have lately said that the atoms of a paramagnetic body have an irrestible urge to be in a state of random orientation whenever there is no applied field acting upon them. The resultant of all the little arrows of a domain should then be zero. How can it have a magnitude which is not merely different from zero, but (on the scale customary for such things) very considerable, and independent of the field strength which is applied to the iron?

The answer to this question is given, and very well given, by that extra field or "Weiss field" within the group, which I first mentioned in connection with the constant θ which paramagnetic solids exhibit. It will be remembered how this constant is explained by assuming that the torque, which acts on any one of the atomic magnets, is due not entirely to the applied field H but to the resultant of that and an extra field AI which is proportional to the magnetization I of the body. We have already had the equation (10) which links I and H when this extra field is present. Now striking H altogether out of that equation, we arrive at this one:

$$I = N\mu \tanh(\mu AI/kT) \qquad (12)$$

which refers to a situation in which there is no applied field at all. This may be regarded as an equation for I, fixing the value or values

of I which can exist in this situation. Now everything which I have said so far encourages the reader to suppose that the only possible value of I in the situation is zero; and as a matter of fact, zero is always a solution of this equation. But suppose that there should be another solution, different from zero. The equation would then assert, that if somehow that value of magnetization should arise in the substance, then the extra field would also arise, and in just the right magnitude to maintain that magnetization perpetually, without any aid in the form of a field applied from the outside.

Well, the equation is not exactly easy to solve for I, but it can be mastered—most conveniently by a graphical way—and the striking result is reached, that if T is greater than θ there is no other solution than $I = 0$, but if T is less than θ there *is* a second solution. I will denote this other by I_w. Consider, then, the situation when there is no applied field: if the temperature is higher than θ, I repeat what I have been saying all along, that random orientation of the atomic magnets is inevitable; but when the temperature is lower than θ, then there is another possibility: there is a stable alignment of the atomic magnets entailing this value I_w of the magnetization, which can maintain itself indefinitely if it should ever come into being. Do not leap to the other extreme, and suppose that this is a *perfect* alignment of the atomic magnets and hence a perfect saturation of the domain. Such a situation could exist (according to the theory) only at absolute zero. The equation gives us I_w as function of T, and this function declines smoothly from the value $N\mu$ (for a domain of unit volume!) at absolute zero, to the value zero at $T = \theta$. The curve between these two points is completely determined by the values of μ and θ, which are derived in such ways as I have indicated from the magnetic properties of the substance at the higher temperatures well above θ.

And now, the culmination. The so-called saturation of iron—the ordinate of the I-vs-H curve when it flattens out and becomes sensibly parallel to the axis of abscissae—is itself (as I mentioned) a function of temperature; *it is this same function* (Fig. 3). What is usually called "saturation" with ferromagnetic bodies consists in aligning the big arrows of the domains, so that in unison of direction they exhibit that value of magnetization which is dictated by their internal temperature and internal field. "True" saturation—"saturation of saturations"—the alignment of the atoms within each domain superposed on the alignment of the domains with the field—this can be attained only at the absolute zero of temperature. We are able, however, to work at temperatures so close to absolute zero, that the remaining degree of extrapolation is slight; and we are able, therefore, to give with much

confidence values for the true saturation of iron, nickel, cobalt, gadolinium, and many ferromagnetic alloys.

(The reader may properly wonder why, instead of solving equation (12) obtained by putting $H = 0$ in equation (10), it is not the practice to put for H the field strengths actually applied to iron when aligning

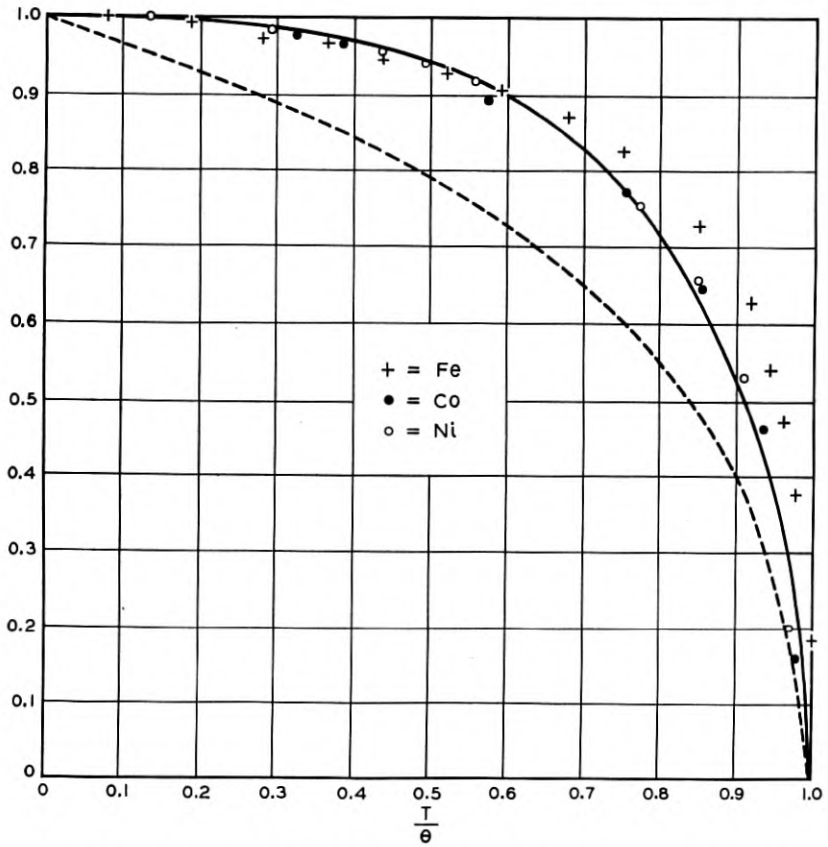


Fig. 3—Intrinsic magnetization plotted against T/θ, for the domains of three ferromagnetic elementary metals (the constant θ has different values for the three). The ordinate is I_w referred to its saturation-value (deduced by extrapolation) as unity. The curves are theoretical, the dashed one by classical theory ($n = \infty$), the full one by quantum-theory ($n = 2$).

the domains or in any other circumstances, and to solve the equation (10) under these conditions? This of course is the correct procedure, but in ferromagnetic bodies AI is usually so enormous by comparison with H, that the latter may be disregarded without appreciable error.)

One naturally asks about the size and the magnetic moment of the domains. It is useless to remember how the latter was determined for paramagnetic bodies from the features of their I-vs-H curves, since the theory which made that possible is not applicable here. Moreover, the super-atoms share with ordinary atoms the quality of being invisible: no feature of the ordinary surface of a metal indicates them, and no technique of etching the surface seems able to delineate them. (It must be said, however, that ferromagnetic powders sprinkled over ferromagnetic metals may distribute themselves in remarkable picturesque patterns, and perhaps these sometimes simulate the pattern of the underlying domains.*) But fortunately the super-atoms are not inaudible; at least, it is not a very extravagant statement to say that they can be heard. Let a girdle of wire around a rod of some ferromagnetic substance be connected through an amplifier with a microphone, and let a gradually-increasing magnetic field act lengthwise on the rod: the microphone will then emit a machine-gun patter of sharp clicks (with suitable amplification it may be very dramatic!) each of which corresponds to the sudden shift of the magnetic moment or "big arrow" of a domain from one of its possible orientations to another. Now if an electrical instead of an acoustical device is attached to the girdle of wire, the magnitude of the moment which thus re-orients itself at a single click may be assessed. It turns out that the moments are of very various magnitudes; a mean may, however, be estimated, and this mean is some 10^{15} times as great as the moment of a single atom. Therefore the average domain comprises a million billions of atoms, and must therefore be about .002 cm in breadth; but there is a wide range of sizes about the average. As for the individual atoms of the ferromagnetic metals, their moments may be derived from equating $N\mu$ to the values (obtained by extrapolation from observations at various low temperatures, to absolute zero) of that "saturation of saturations" defined above. They are by no means out of the common. Iron and its congeners are readily magnetizable, not because their atoms are extraordinarily magnetic—which is not at all the case—but because their atoms have this curious propensity of cohering together in large groups, developed to an extraordinary degree.

To many features of ferromagnetism, of which whole monographs might be or have been written, I can give only brief mention or none at all. There are the "magneto-caloric effects," arising because, when a ferromagnetic body is heated, the dis-alignment of the atoms

* Cf. the article of R. M. Bozorth in the preceding number (January 1936) of this Journal.

within each domain increases, and this increase requires additional heat over and above that which goes to augment the kinetic energy of the atoms. The specific heat of iron (as of its congeners) is greater than it would be, but for this effect; the excess may be computed from the foregoing theory as function of temperature, and the computed values agree with the data to an extent which speaks very strongly for the theory. (The like is the case with a paramagnetic body exposed to a magnetic field; and as a result, such a body will grow cooler when the applied field is withdrawn, the kinetic energy of the atoms being levied upon when the dis-alignment occurs. The effect is imperceptible in usual circumstances, but with such substances as iron-ammonium alum at liquid-helium temperatures, it becomes so strong that the lowest temperatures ever achieved have been attained by making use of it.) There are the "magnetostrictive effects," arising because, when the atoms of the domains change their orientation, the metal as a whole is strained. It follows that there are interrelations between magnetization, strain, and stress; and anyone remembering even a little of the mathematical theory of elasticity with its moduli and its stress-strain tensors will readily believe that the theory of these interrelations is marvelously complicated. As one sensational example of the consequences, I cite the fact that when a certain permalloy is exposed to a field of, say, one half of one gauss, its magnetization ranges between a few per cent and nearly one hundred per cent of saturation, according to the strength of the tensile stress applied to it. The many processes of the metallurgical arts have often vast effects upon the magnetic properties of the ferromagnetic metals exposed to them: some are due to the changes in the elasticity and hence in the magnetostrictive effects, some to the changes in the chemical constitution (e.g. in the proportion of impurities), some to the changes in phase (of alloys) which these processes entail; but it would be risky to affirm that they have all been traced to one or another of these causes. The finer details in the shape of the I-vs-H curve for ferromagnetics remain to be explained, and to account for one of them it seems to be thought necessary to assume that the domains may gain or lose in size at one another's expense; it is too bad that this impairs the concept of the domain as an immutable super-atom. I leave without overmuch regret this infinitely detailed and complicated topic, to conclude by brief allusions to the spinning electron and to diamagnetism.

Hitherto in these pages I have let it be inferred that when we obtain the magnetic moment of the atom of some element or the molecules of some compound by magnetic experiments upon the substance, it

always agrees with the "theoretical" value derived from the spectrum of that substance when a gas. This is indeed the case with gases and even with a certain number of solids, a large enough number to inspire confidence in the theory. There are, however, numerous exceptions among solids—a circumstance not to be wondered at, since an atom incorporated in a solid is usually in a very different condition from an atom freely wandering about in a gas. The like is true about that number n, the "number of permitted orientations of the atom in a field," which was introduced near the beginning of the article. Either the trend of the I-vs-H curve for a paramagnetic, or the trend of the I_w-vs-H curve for a ferromagnetic, enables us (if it has been sufficiently well measured) to ascertain the value of n; and in a surprising number of instances, comprising iron, cobalt and nickel as well as various rare-earth elements in chemical compounds, the curves prescribe the value *two*, when the free atom according to its spectrum would display some other value. Thus when the atoms are compacted together into a solid, their proximity affects them in such a way as to bring about this result.

Now the important point about this value *two* for n is, that it is the value to be expected for an electron which is either isolated, or else linked to its atom in such a way that it has no orbital angular momentum. The contemporary theory of spectra includes, as one of its essential elements, the postulate of the "spinning electron"—the postulate that each electron by itself is endowed with an intrinsic and indestructible angular momentum and magnetic moment, of definite known amounts, having nothing whatever to do with its orbital revolutions.[4] This angular momentum or "electron-spin" is of the amount which requires $n = 2$, when it is not compounded with an angular momentum of orbital motion or with angular momenta of other electrons. The atoms in question behave, when compacted into solids, as though this angular momentum of individual spinning electrons were the only one left outstanding.

This striking inference is greatly strengthened by measurements upon the one phenomenon in which that angular momentum, which according to atomic theory is always the companion of magnetic moment, comes to light. Imagine a cylinder of some paramagnetic or ferromagnetic substance, hanging freely from a suspension attached to one end. Suppose it to be unmagnetized at first; this signifies that the atoms (whether or not they are grouped into domains) are so oriented that the resultant of all their angular momenta, as well as

[4] The reasons furnished by spectroscopy for making this postulate are much too complex to be interpolated in this article: I refer to the first fourteen pages of "Contemporary Advances in Physics," XXIX, this *Journal*, 14, 285–321 (April 1935).

that of all their magnetic moments, are zero. Now suppose that a field is suddenly applied, parallel to the axis of the cylinder. The substance is suddenly magnetized; this signifies that the resultant of the magnetic moments, and hence that of the angular momenta, are no longer zero. Let I stand (as heretofore) for the former resultant, and P for the latter. Now it is desirable to remember that each atom consists of a nucleus and an electron-family; that the electron-family possesses the magnetic moment and is oriented in the field (it strictly is what I have hitherto referred to as "the atom"); that the nuclei of the atoms in the cylinder are relatively non-magnetic but contain nearly all of the mass of the cylinder.[5] At the moment of magnetization, the ensemble of the electron-families acquires a net angular momentum P. Now angular momentum being one of these things (like energy and linear momentum) of which the total in Nature does not vary, an equal and opposite amount, $-P$, must appear somewhere or other. It appears in the mass of the cylinder, presumably because of some interaction between the electron-families and their nuclei. The cylinder makes a sharp turn at the instant of magnetization, twisting the suspension from which it hangs through an angle from which (and from the rigidity of the suspension) the value of $-P$ can be found. This effect and its converse (an unmagnetized cylinder may be magnetized by sharply twisting it) are known as the "gyromagnetic effects." They are delicate and difficult to produce, *a fortiori* to measure; yet of late years experimenters have succeeded in measuring P together with I, and therefore learning the value of the ratio I/P—first for the ferromagnetic metals and then for some of their compounds and alloys, and lately for certain paramagnetic salts, the work on these last being done at the very low temperatures where alone they can be strongly magnetized.

This ratio I/P—its reciprocal is called the "gyromagnetic ratio"—is a rare sort of thing: it is a quantity of which the numerical value, measured on pieces of bulk matter, is appropriate also to the elementary particles. If the substance is made up of identical elementary magnets of magnetic moment μ and angular momentum p, then I/P is μ/p. Since μ and p are knowable from spectra, so also is their ratio. Its lowest possible value (from theory) is $e/2mc$, in which e, m, and c have their usual meanings;[6] this would always occur if the electrons had no spins; actually it occurs if the electron-family of the atom is so

[5] Most nuclei possess magnetic moments, which, however, are so excessively small that they can be detected only by experiments of extreme delicacy.

[6] Charge (in E.S.U.) and mass of the electron, and speed of light in vacuo. For the theory underlying these statements, c.f. *l.c.* pp. 285–300. Often the ratio of the experimental value of μ/p to the quantity $e/2mc$ is called an "experimental g-value," the ratio of the theoretical value to $e/2mc$ being conventionally denoted by g.

organized that the spins neutralize one another. Its highest possible value is twice as great; this occurs if nothing counts excepting the electron-spins, and signifies either that the electrons are free [7] or else that the electron-family of each atom is so organized that there is no net angular momentum due to orbital motion. Intermediate values are possible and signify different types of organization of the electron-family. The values predicted from spectra have been confirmed for a few of the rare-earth atoms in their paramagnetic salts; but usually, as I have already intimated, the *observed value of the ratio* μ/p is about $2(e/2mc)$, though the spectrum says something else.

It would be pleasant now to add that the magnetic moment of each of these substances, per atom, amounts to some integer multiple of the magnetic moment μ_e of the spinning electron. We then could say that the integer is the number of "uncompensated" spinning electrons in the atom, implying by the word "uncompensated" in this connection that all the magnetic moments in the electron-family of the atom add up vectorially to zero and so do all the angular momenta, with the sole exception of those pertaining to these electron-spins. Such is not, however, the case: some of the experimental values are $2.2\mu_e$ for iron, $1.7\mu_e$ for cobalt, $0.6\mu_e$ for nickel. It seems necessary to assume that in metallic solid iron, some of the atoms present two uncompensated electrons to the orienting field, and others three. Iron in different chemical compounds exhibits different values of magnetic moment, and sometimes the ratio μ/p is different from $2(e/2mc)$, suggesting that angular momenta of orbital motion are not quite cancelled out; indeed it now appears that the ratio is slightly but definitely different from this specific value even in the cases (such as those of the pure ferromagnetic metals and of permalloy) in which at first the measurements suggested that it was the same.

Such observations as these last are problems for the specialists in atomic theory; magnetism offers great numbers of these problems. Another and a complementary way of viewing this situation is, to look on every measurement of a magnetic moment made upon a solid as an item of information about an atom (or a molecule) existing in a condition which is not accessible to spectroscopic research. Spectra indicate the normal state of atoms in freedom; occasional magnetic experiments (like those on gaseous oxygen here cited, or those on molecular beams by the Gerlach-Stern method, which I hope to treat on a later occasion) also refer to free atoms and molecules, and confirm the indications of the spectra, thus sustaining both the methods; but

[7] Certain metals, the alkali metals for instance, exhibit a paramagnetism which is entirely due to the "free" or conduction electrons.

mostly the magnetic methods refer to atoms in a solid, and so they make available a new and broad domain for the operations of atomic theory.

There remains diamagnetism. The first thing to be said about the theory of diamagnetism is discouraging; for it has the earmark of a futile atomic theory—it involves the assumption that the individual atoms behave exactly like the substance as a whole. Under all field strengths and all conditions, it is assumed that the diamagnetic moment of a block of N atoms is N times the diamagnetic moment of a single atom. Nevertheless this is not a futile assumption, for strictly it is not an assumption at all but an inference from atomic structure. It was mentioned early in these pages that owing to the unbreakable link between angular momentum and magnetic moment, a magnetic atom precesses about the direction of the field. This motion of precession is an extra motion of the electrons of the atoms, a circulatory motion around the axis supplied by the direction of the field. This extra motion entails an extra current, which entails an extra magnetic moment, which is the source of diamagnetism or which *is* diamagnetism. *Diamagnetism is precession.* It is not confined, as the foregoing words suggest, to atoms which have a net magnetic moment. Consider an atom (a free atom of any noble gas will afford an example) possessing two or more electrons, the orbits and the spins of which are so oriented that the resultant magnetic moment is nil. Though some of the orbits and spins are pointed oppositely to others, they all precess in the same sense, and the atom acquires a magnetic moment in the field though it had none beforehand. The like is true, of course, when the resultant of the orbits and the spins is different from zero; the agents of orientation which were discussed above render it paramagnetic, but the precession renders it diamagnetic, and it is paramagnetic and diamagnetic—or ferromagnetic and diamagnetic—at one and the same time. The moment due to the precession is proportional to the field strength, and the factor of proportionality may be calculated from the structure of the atom (it depends primarily upon the areas of the electron-orbits). The agreement of the calculated values with the data is generally satisfactory; and diamagnetism, the least conspicuous of the three types of magnetism, takes precedence over the others as being that one of the three of which our understanding is most nearly perfect.

The Proportioning of Shielded Circuits for Minimum High-Frequency Attenuation

By E. I. GREEN, F. A. LEIBE and H. E. CURTIS

For given conditions of design there exists an optimum proportioning or configuration which makes the high-frequency attenuation of a given type of individually shielded circuit a minimum. Determination is made of such optimum proportioning for a wide variety of types of individually shielded circuits including several novel types designed to make the high-frequency attenuation low in comparison with the cross-sectional area occupied by the circuit, and the attenuation of different types is compared. The following topics and specific circuit structures are considered:

COAXIAL CIRCUITS—Basic Coaxial Circuit; Effect of Dielectric; Effect of Frequency on Optimum Ratio; Thin Walls; Stranded Conductors; Optimum Proportioning as a Function of Conductor Resistance.

BALANCED SHIELDED CIRCUITS—Shielded Pair (Cylindrical Conductors and Shield)—Condition for Minimum Attenuation, Condition for Maximum Characteristic Impedance, Effect of Dielectric, Effect of Frequency; Pair in Space; Shielded Stranded Pair; Pair with Shield Return; Double Coaxial Circuit; Shielded Pair (Round Conductors and Oval Shield); Shielded Pair (Quasi-Elliptical Conductors); Shielded Quad.

INTRODUCTION

SINCE the very beginning of mathematics, problems of maximizing and minimizing have possessed a marked fascination. The Greeks were successful in solving a few geometric problems of this character. Later, algebra was found to be another method of attack. Finally, the powerful methods of the calculus became available for the determination of maxima and minima in manifold variety. The reasons for the continued interest in such problems are not hard to find. It is but natural to seek the ideal, and here, at least, is one phase of mankind's search for perfection in which a goodly measure of success may be achieved. In addition, a knowledge of the optimum dimensioning of things, or of the optimum relations between things, frequently holds much practical significance.

It is mainly with problems of maxima and minima that this paper is concerned. These problems have to do with transmission circuits which are surrounded by individual shields. Recent literature [1, 2] has pointed out that circuits of this type have properties which render them especially suitable for the transmission of broad bands of frequencies. Such circuits are also finding application as "lead-ins" to connect radio antennas with transmitting or receiving apparatus.[3, 4]

[1] For numbered references, see end of paper.

It is well at this juncture to understand the function of shielding in a high-frequency transmission circuit. Such shielding serves one or both of these purposes: (*a*) keeping interference due to external sources from entering the circuit, and (*b*) preventing the circuit from causing interference in external circuits. The shielding may either supplement or completely replace the use of electrical balance to reduce interference. The design of shield, that is, its construction, material, thickness, etc., is determined by the degree of shielding required and by considerations of mechanical performance and cost. The degree of shielding needed depends in turn upon such factors as the type and length of circuit, the nature and frequency of the signals to be transmitted, and the magnitudes of external interference. These interesting aspects of shield design, some of which have been dealt with elsewhere,[1, 2, 5] will not be discussed here.

Attention will rather be directed to an intriguing property of any individually shielded circuit, namely, that, for given conditions of design, there always exists an optimum proportioning or configuration which makes the transmission efficiency of the circuit a maximum, or, in other words, makes the attenuation a minimum. One such condition of design which may be imposed is that the cross-sectional area enclosed within the shield is to be a constant. In what follows, determination will be made of such optimum proportioning for a wide variety of types of individually shielded circuits. Since the attenuation is generally of outstanding importance in a high-frequency transmission line, the results should be not only of theoretical interest but also of practical value. Moreover, the different methods which are used in solving these problems should find further application, both in the many other known problems which must perforce be omitted for lack of space, and in those problems which may be conceived in the future.

The principal types of individually shielded circuits to be discussed are:

(1) Coaxial or concentric circuits, in which an outer conductor, which serves also as a shield, completely surrounds a centrally disposed inner conductor.
(2) Shielded pairs, consisting of a pair of conductors which form the transmission circuit, these being surrounded by an individual conducting shield.

The coaxial circuit is unbalanced, and relies solely upon shielding for protection against interference from or into its exterior. In contrast to this is the balanced type of circuit, in which the go and return

conductors are designed to be substantially alike and are located substantially symmetrically with respect to earth and surrounding conductors.

In the past, telephone transmission circuits have been largely of the balanced type. It has been found possible to operate such balanced circuits up to fairly high frequencies,[2] without incurring excessive interference. However, as the frequency is raised it becomes increasingly difficult to maintain a sufficiently high degree of balance, and shielding may then be desirable. The shielding may eliminate balance entirely, as in the coaxial circuit, or may be combined with balance in what may be termed a shielded balanced circuit, of which the shielded pair is an outstanding example.

For the simplest forms of circuits, the optimum relations may be precisely derived with the aid of the propagation formulas. In more difficult cases it is necessary to use approximate methods of one kind or another. These methods, however, can generally be made to yield sufficiently accurate results for practical purposes.

Coaxial Circuits

Coaxial circuits, which furnish the least difficult problems in optimum proportioning, make a natural starting point for this subject.

Basic Coaxial Circuit

The first type of circuit to be considered is the basic circuit consisting of two tubular conductors arranged coaxially, whose cross-section is shown diagrammatically in Fig. 1.

Before trying to find out how to proportion such a circuit, it must be noted that in the design of any shielded circuit there enter a number of variables, including the overall size of the structure, the type and thickness of shield, the type of conductor or conductors, the type of insulation, and the frequencies to be transmitted. Some of these factors exert an important influence on the optimum proportioning, so that it is necessary, in order to arrive at a unique solution in a given case, to keep certain factors fixed. Thereafter, however, the effect produced upon the result by varying these factors may be examined.

First, therefore, let the following assumptions be made:

1. That the tubular conductors of Fig. 1 are composed of solid material.
2. That the dielectric is gaseous, with zero dielectric loss. This is a condition which may be approached in practice.
3. That the inner diameter of the outer conductor is fixed. This is a convenient assumption, having for its basis the fact that it is ordinarily desirable, for economic or other reasons, to limit the

size of the outer conductor, and the further fact that the thickness of the outer conductor will ordinarily be determined by mechanical considerations or by shielding requirements.

4. That the frequency is high enough to permit the use of certain approximate formulas as noted below. Practically, this means that at the frequency considered the currents are largely crowded toward the inner surface of the outer conductor and the outer surface of the inner conductor.

The problem is to discover the proportioning which will make the high-frequency attenuation of the circuit a minimum under such conditions. It is well known that the attenuation of a transmission

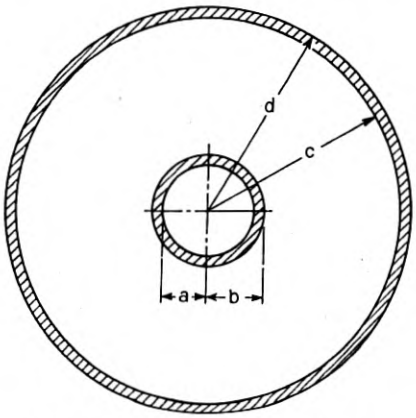


Fig. 1—Coaxial conductor circuit.

circuit at high frequencies may be represented by the following approximate formula:[6]

$$\alpha = \frac{R}{2}\sqrt{\frac{C}{L}} + \frac{G}{2}\sqrt{\frac{L}{C}} \text{ nepers per cm.,} \qquad (1)$$

where R, L, G and C designate, respectively, the linear resistance, inductance, conductance and capacitance of the circuit. Except as otherwise indicated, values in this and subsequent formulas are expressed in c.g.s. electromagnetic units.

When the dielectric loss is negligible, the second term of formula (1) evidently disappears.

Let a and b represent, respectively, the inner and outer radii of the inner conductor, c and d the inner and outer radii of the outer conductor, f the frequency, λ_1 and μ_1, respectively, the conductivity and

permeability of the material of the inner conductor, and λ_2 and μ_2 the corresponding values for the outer conductor. The ratio λ_1/λ_2 will be designated by n.

The high-frequency resistance of the inner conductor may then be approximately expressed by the formula:[5, 7]

$$R_1 = \frac{1}{b}\sqrt{\frac{f\mu_1}{\lambda_1}} \text{ abohms per cm.} \tag{2}$$

Similarly the high-frequency resistance of the outer conductor is approximately:

$$R_0 = \frac{1}{c}\sqrt{\frac{f\mu_2}{\lambda_2}} \text{ abohms per cm.} \tag{3}$$

The high-frequency inductance of the circuit is approximately [7]

$$L = 2 \log_e \frac{c}{b} \text{ abhenries per cm.} \tag{4}$$

The capacitance of the circuit is [8]

$$C = \frac{\epsilon}{2 \log_e \frac{c}{b}} \text{ abfarads per cm.,} \tag{5}$$

where ϵ is the dielectric constant of the dielectric material between conductors, equal to $1/9 \times 10^{-20}$ for gaseous dielectric, corresponding to unity in the practical system of units.

The high-frequency attenuation of the coaxial circuit with negligible dielectric loss, obtained by combining the above formulas, is

$$\alpha = \frac{1}{2c}\sqrt{\frac{f}{\lambda_1}}\left(\frac{c}{b} + \sqrt{n}\right)\frac{\sqrt{\epsilon}}{2 \log_e \frac{c}{b}} \text{ nepers per cm.} \tag{6}$$

The value of permeability assumed in the above equation, and hereafter, is unity, but the methods may be used also for other values.

If the inner diameter of outer conductor be assumed fixed, this expression may be minimized with respect to the ratio c/b, which is the ratio of the radii (or diameters). For convenience this ratio may be designated as ρ. It is found that the high-frequency attenuation is a minimum when the value of ρ is that given by

$$\log_e \rho = \frac{\rho + \sqrt{n}}{\rho}. \tag{7}$$

Figure 2 shows the values of the ratio ρ which satisfy this relation plotted as a function of the conductivity ratio n.

It is noteworthy that the optimum ratio of radii or diameters is independent of (*a*) the diameter and thickness of outer conductor, (*b*) the inner diameter of the inner conductor, and (*c*) the frequency, provided the frequency is high enough for the approximate formulas to hold. It follows from (*a*) that, assuming a fixed thickness of outer conductor, moderately small in comparison with its diameter, relation (7) makes it possible to find the minimum size of outer conductor with which a given value of high-frequency attenuation may be realized. It follows from (*b*) that the inner conductor may be either hollow or solid, provided that the approximate resistance formulas are valid.

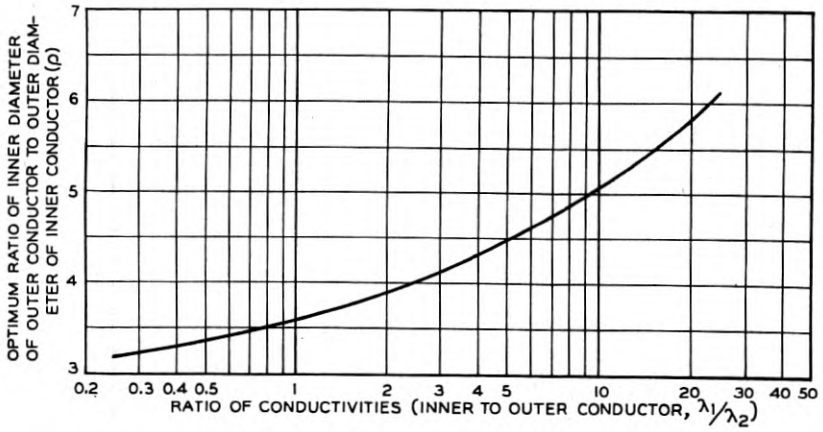


Fig. 2—Variation of optimum diameter ratio of coaxial circuit with conductivity ratio.

A case of special interest arises when the two conductors have the same conductivity, that is, when n equals unity. For this condition the solution of (7) is *

$$\rho = \frac{c}{b} = 3.59. \tag{8}$$

A practical example of the case of different conductivities is a coaxial structure in which the inner conductor is of copper and the outer conductor of lead. For a lead outer conductor containing about 1 per cent of antimony, the ratio of conductivities of inner and outer

* The existence of an optimum relation of this kind was first noted by C. S. Franklin, who gave the value as 3.7. (See Reference 3.) Subsequently the precise value was derived independently of Franklin. (See Reference 10.)

conductors is approximately 13, and the optimum diameter ratio for such a structure, as found from Fig. 2, is about 5.25.

The behavior of the attenuation in the vicinity of the optimum diameter ratio is illustrated in Fig. 3, which shows attenuation plotted

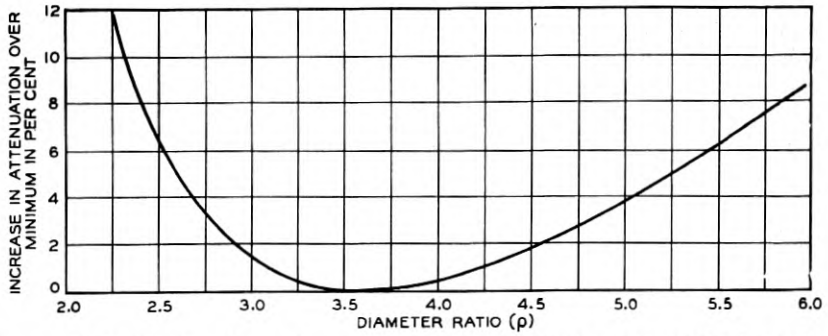


Fig. 3—Variation of optimum diameter ratio of coaxial circuit with conductivity ratio.

against diameter ratio for the case where n equals unity. It will be seen that near the optimum the attenuation changes very slowly. This is fortunate, since it means that unavoidable departures from the optimum diameter ratio may be permitted without appreciable effect on the attenuation. Other small departures from ideal design are also allowable. Thus, for example, it has been assumed in deriving the condition for minimum attenuation that the two conductors of the circuit are perfectly coaxial or concentric. However, for moderately small departures from perfect concentricity occasioned by practical difficulties of construction, the conditions for minimum attenuation are substantially the same as for a circuit with no eccentricity. The situation is similar for other types of shielded circuits to be considered later, in these cases also only a reasonably close approximation to the ideal being necessary.

Effect of Dielectric

Suppose now that the capacitance and leakage conductance introduced by the insulation are substantial.[9] First, it will be assumed that the space between the two conductors is filled with a substantially uniform non-gaseous dielectric material having a dielectric constant ϵ and a power factor p. Such would be the case, for example, if the two coaxial conductors were separated by a continuous rubber insulation. The leakage conductance of the circuit now becomes

$$G = p\omega C = \frac{p\omega\epsilon}{2 \log_e \frac{c}{b}} \text{ abmhos per cm.,} \tag{9}$$

where, as usual, ω equals $2\pi f$.

By substituting in formula (1), the high-frequency attenuation is found to be

$$\alpha = \frac{1}{2c}\sqrt{\frac{f\epsilon}{\lambda_1}}(\rho + \sqrt{n})\frac{1}{2\log_e \rho} + \frac{p\omega\sqrt{\epsilon}}{2} \text{ nepers per cm.} \qquad (10)$$

Since ω, p, and ϵ are not functions of the ratio c/b, the second term of this expression is constant for purposes of differentiation with respect to that ratio, and the condition for minimum attenuation is identical with that previously found, as given in formula (7).

A high-frequency transmission property of smaller interest than the attenuation is the characteristic impedance. This is given by the familiar formula [6]

$$Z_0 = \sqrt{\frac{L}{C}} \text{ abohms.} \qquad (11)$$

For the coaxial circuit with dielectric constant ϵ the high-frequency characteristic impedance is

$$Z_0 = \frac{2\log_e \rho}{\sqrt{\epsilon}} \text{ abohms.} \qquad (12)$$

There now comes the case where the space between the conductors consists of a combination of gaseous and non-gaseous dielectrics. Perhaps the simplest example occurs when the conductors are separated by insulating discs or washers extending continuously between the two conductors with flat sides perpendicular thereto. Such a construction is illustrated in Fig. 4. Let the thickness of each insulating

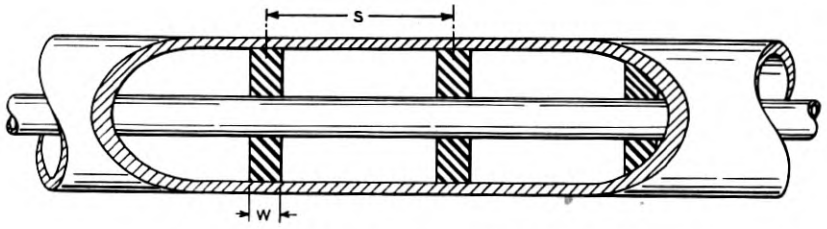


Fig. 4—Coaxial structure with disc insulation.

disc be designated w, the spacing between centers of adjacent discs, s, the dielectric constant of the air dielectric, ϵ_1, and that of the disc material ϵ_2.

The capacitance of the coaxial circuit now becomes

$$C = \frac{\epsilon_1 + \frac{(\epsilon_2 - \epsilon_1)w}{s}}{2 \log_e \rho} \text{ abfarads per cm.,} \tag{13}$$

while the leakage conductance is

$$G = \frac{w}{s} \cdot \frac{p\omega\epsilon_2}{2 \log_e \rho} \text{ abmhos per cm.} \tag{14}$$

On substituting these values in formula (1) the following expression results:

$$\alpha = \frac{1}{2c}\sqrt{\frac{f}{\lambda_1}}\frac{\sqrt{\epsilon_1 s + (\epsilon_2 - \epsilon_1)w}}{\sqrt{s}}\frac{\rho + \sqrt{n}}{2 \log_e \rho} + \frac{p\omega\epsilon_2}{2}\frac{w}{\sqrt{s}}\frac{1}{\sqrt{\epsilon_1 s + (\epsilon_2 - \epsilon_1)w}} \text{ nepers per cm.} \tag{15}$$

Once more the second term is independent of c and b, and the condition for minimum attenuation is, as before, that given by equation (7).

The high-frequency characteristic impedance in this case, however, is

$$Z_0 = \frac{2 \log_e \rho}{\sqrt{\epsilon_1 + \frac{(\epsilon_2 - \epsilon_1)w}{s}}} \text{ abohms.} \tag{16}$$

The quantity in the denominator of the above expression is evidently the weighted average dielectric constant of the insulating medium.

In the case just considered, the gaseous and non-gaseous dielectrics were separated from each other by planes perpendicular to the axis of the conductors. Consequently, each line of dielectric flux passed through only one kind of material. It can be shown that, as long as this latter condition holds, the condition for minimum high-frequency attenuation as given by equation (7) is valid, or, in other words, the optimum diameter ratio is that shown in Fig. 2. Cases arise, however, in which a line of dielectric flux, in going from one conductor to the other, may pass through more than one kind of dielectric material. It is extremely difficult to obtain a mathematical solution for the diameter ratio which results in minimum attenuation for such cases, since this involves a three-dimensional field problem. Consideration of the problem, however, indicates that the optimum diameter ratio will not differ appreciably from that given by Fig. 2, especially if the dielectric is mostly gaseous, which, of course, is highly desirable.

Effect of Frequency on Optimum Ratio

It has been seen that, at the higher frequencies where the approximate transmission formulas may be employed, the optimum diameter ratio is substantially independent of frequency. In so far as the practical application of individually shielded circuits is concerned, it is in these higher frequencies that interest primarily centers. Even when it is desired to transmit a wide band extending from high frequencies down to comparatively low ones, it is advantageous to proportion the circuit so as to minimize the attenuation at the highest transmitted frequency, since the attenuation at all lower frequencies will be less than the value thus obtained.

It may, however, be worth while to consider briefly the question of optimum proportioning when low frequencies only are involved. The appropriate transmission formulas to be used instead of the approximate high-frequency expressions are known,[5] and the optimum diameter ratio in any specific case may be derived from these. It will be evident that, since skin effect is present to a lesser degree at the low frequencies, the diameter and thickness of the outer conductor and the thickness of the inner conductor will, as the frequency is decreased, have an increasing influence on the optimum proportioning.

Without attempting to derive precise values for the different conditions, it may be noted that the optimum diameter ratio for low frequencies is invariably less than that for high frequencies, the high-frequency value being approached asymptotically as a limit. The reason for this will be readily apparent. Let the inner diameter and thickness of the outer conductor be assumed fixed. At high frequencies the resistance of the inner conductor varies inversely with the first power of its diameter. At lower frequencies, however, this resistance varies inversely with some power of the diameter greater than unity, and finally, at zero frequency, assuming a solid wire, with the square of the diameter. Hence it is, that, in varying the size of the inner conductor in order to obtain a balance between the change of resistance and change of capacity, it is advantageous to make the inner conductor somewhat larger, or, in other words, to make the diameter ratio smaller, at low frequencies than at high frequencies.

Thin Walls

What is the result if the walls of the two coaxial conductors are made very thin? Under this condition the conductor resistance, and hence the attenuation, will remain substantially constant over a wide range of frequencies. This constancy is realized, however, at the expense of an increase in the attenuation as compared with that for thicker conductor walls.

Using the notation of Fig. 1, the resistances of the inner and outer conductors, both with conductivity λ, at frequencies where the walls are sufficiently thin to avoid skin effect, are

$$R_i = \frac{1}{\pi\lambda(b^2 - a^2)} \text{ abohms per cm.} \tag{17}$$

$$R_0 = \frac{1}{\pi\lambda(d^2 - c^2)} \text{ abohms per cm.} \tag{18}$$

Let the inner conductor have a fixed thickness $b - a$, the outer conductor a thickness $d - c$, and let the ratio $(b - a)/(d - c)$ be represented by t. For small values of wall thickness

$$b^2 - a^2 \doteq 2b(b - a) \tag{19}$$

and

$$d^2 - c^2 \doteq 2c(d - c) = 2c\frac{(b - a)}{t}. \tag{20}$$

Substituting these relations and the values of L and C from (4) and (5) in equation (1), it is found that the attenuation for the circuit with thin walls is

$$\alpha = \frac{\sqrt{\epsilon}}{4\pi\lambda c}\frac{(\rho + t)}{(b - a)}\frac{1}{2\log_e \rho} \text{ nepers per cm.} \tag{21}$$

Differentiation shows that minimum attenuation in the case of thin walls is obtained when

$$\log_e \rho = \frac{\rho + t}{\rho}. \tag{22}$$

The values of diameter ratio which satisfy this relation may be found from the curve of Fig. 2, if the values of abscissæ on that curve are interpreted as values of t^2.

If the conductor walls are thin, as above, and if in addition the conductivities of the two conductors are not the same, that of the inner conductor being n times that of the outer one, the condition for minimum attenuation becomes

$$\log_e \rho = \frac{\rho + nt}{\rho}. \tag{23}$$

Figure 2 may be used to find the values of diameter ratio which satisfy this relation also, the abscissæ scale markings in this case being taken as values of n^2t^2.

Stranded Conductors

With conductors having solid walls, or composed of non-insulated strips or filaments, the currents at high frequencies are largely crowded toward the inner surface of the outer conductor and the outer surface of the inner conductor, due to skin effect. Since the losses in the conductors themselves ordinarily comprise the major portion of the attenuation in a coaxial circuit, interest attaches to the possibility of counteracting the increase in conductor resistance due to skin effect by using a conductor composed of a number of individually insulated strands so twisted or interwoven as to distribute the current more nearly uniformly over the cross-section.[11] Chief attention naturally focuses upon the inner conductor, which is by far the greater contributor to the resistance, and this discussion will be largely limited to the case where only the inner coaxial conductor is stranded.*

Types of stranded conductors suitable for use as the inner conductor of a coaxial circuit include both those in which the conductor cross-section is completely filled with insulated strands and those in which the insulated strands form an annular cross-section, surrounding a core of non-conducting or conducting material. Of various possible methods of stranding, one simple and effective process is similar to that used in the construction of rope. A few strands are twisted together to form a group, several such groups are twisted into a larger group, and so on until the desired conductor cross-section is obtained.

The high-frequency resistance of a stranded conductor may be determined either by measurement or computation. For a completely stranded inner conductor of any diameter, size, number of strands, and thickness of insulation, the high-frequency resistance is given by S. Butterworth [12] and in unpublished material by J. R. Carson. The resistance values obtained in measurements of stranded conductors approximate very closely the theoretical results.

In evaluating the results obtained with stranding, it is convenient to compare the resistance of a stranded conductor with that of a non-stranded conductor of the same overall size. For the case of a stranded inner conductor, the ratio of the resistance of the stranded conductor at any given frequency to the resistance at the same frequency of a solid conductor having the same outer diameter and composed of the same material used in the strands may be designated as m.

The values of the resistance ratio m which may be realized in practice depend upon the frequency and the design of stranded conductor. Some idea of these values for two specific conductors may be obtained

* "Stranded" is used to mean "composed of insulated strands."

from the curves of Fig. 5. It will be seen that there is ordinarily a frequency at which the resistance ratio is a minimum. Above this frequency the improvement due to stranding rapidly vanishes, the performance thereafter being worse than that of the corresponding non-stranded conductor. The minimum value of resistance ratio attained in the range of some hundreds of kilocycles may be in the order of 0.6, a very substantial improvement. In order to secure any marked advantage in the frequency range above 700 or 800 kilocycles, the number and fineness of the individual strands would be such as practically to preclude their use.

Another result obtained with stranding is an increase in the internal inductance of the conductors, which likewise serves to reduce the high-

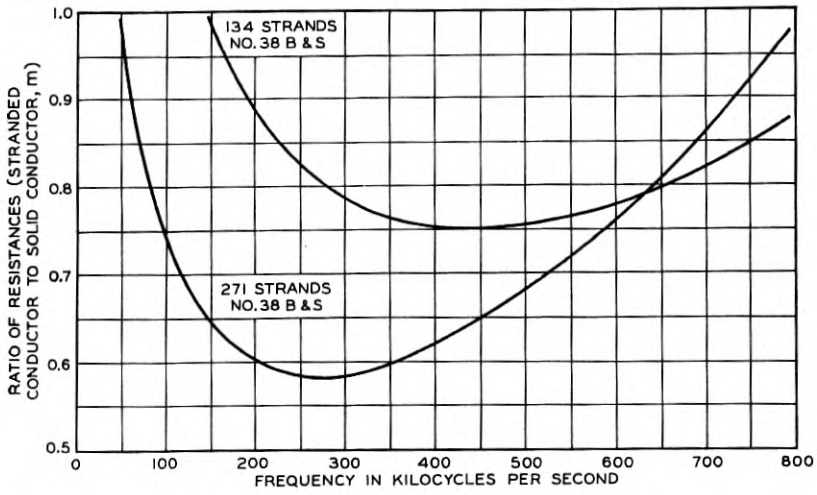


Fig. 5—Resistance ratios of stranded conductors.

frequency attenuation. For a round conductor which is completely stranded, the internal inductance at all frequencies where the current is uniformly distributed over the conductor cross-section approximates .5 abhenry per centimeter, which is the internal inductance of a solid round wire at zero frequency. In general, this value of internal inductance will hold up to frequencies somewhat above that for which the resistance ratio m is a minimum. The internal inductance of a stranded conductor of annular cross-section, for all frequencies where the current is uniformly distributed over the cross-section, is

$$L_i = \frac{b^2 - 3a^2}{2(b^2 - a^2)} + \frac{2a^4}{(b^2 - a^2)^2} \log_e \frac{b}{a} \text{ abhenries per cm.} \qquad (24)$$

This is the same as the internal inductance at zero frequency of a solid tube of the same dimensions.

Since either the inner or outer conductor of a coaxial circuit, or both, may be stranded, and since, in addition, the dielectric loss may either be negligible or may be appreciable, there are six different cases of optimum proportioning which might be considered.[13] Only one case, however, that of a coaxial circuit with only the inner conductor stranded and with negligible dielectric loss, will be taken up here. The high-frequency attenuation of such a coaxial circuit is

$$\alpha = \frac{m}{2c}\sqrt{\frac{f}{\lambda_1}}\left(\rho + \frac{\sqrt{n}}{m}\right)\sqrt{\frac{\epsilon}{(4\log_e \rho)^2 + 2L_i \log_e \rho}} \text{ nepers per cm.} \quad (25)$$

While the value of m varies with frequency and with the design of the stranded conductor, this value is, for a particular frequency and a particular design, definitely determinable. As has been noted, it is generally desirable to proportion a transmission circuit so as to minimize the attenuation at the highest frequency to be transmitted. Furthermore, the value of m will not vary rapidly with changes in conductor diameter provided the number of strands be changed as the conductor size is varied. It therefore becomes possible to treat m as a constant in deriving the relation for optimum proportioning.

Using ρ to designate c/b, the condition for minimum high-frequency attenuation is found to be

$$\frac{2\frac{m}{\sqrt{n}}\rho \log_e \rho}{\frac{m}{\sqrt{n}}\rho + 1} = \frac{4\log_e \rho + L_1}{2\log_e \rho + L_1}. \quad (26)$$

Figure 6 shows graphs of equation (26) for two values of L_i, namely, $L_i = 0.5$ abhenry per centimeter, which corresponds to the case where the cross-section of the inner conductor is completely stranded, and $L_i = 0$. When the stranded inner conductor is of annular cross-section the optimum value of the diameter ratio lies somewhere between the two curves shown. The useful range of m probably lies between about 0.5 and unity and that of n between about 1 and 15.

As to the practical use of stranding, it is apparent from the resistance ratio curves of Fig. 5 that in order to take advantage of stranding it would be necessary to limit the transmission band to a maximum frequency well below that possible with non-stranded conductors. Further drawbacks to the use of stranded conductors are their greater cost as compared with non-stranded ones, and greater mechanical

difficulties in using them. For these reasons stranded conductors do not seem likely to find early application in broad band transmission circuits.

Optimum Proportioning as a Function of Conductor Resistance

The optimum diameter ratio of a coaxial circuit may also be expressed broadly as a function of the two conductor resistances. Assume a coaxial circuit in which the high-frequency resistance of the inner conductor varies, at least over a limited range, inversely as its

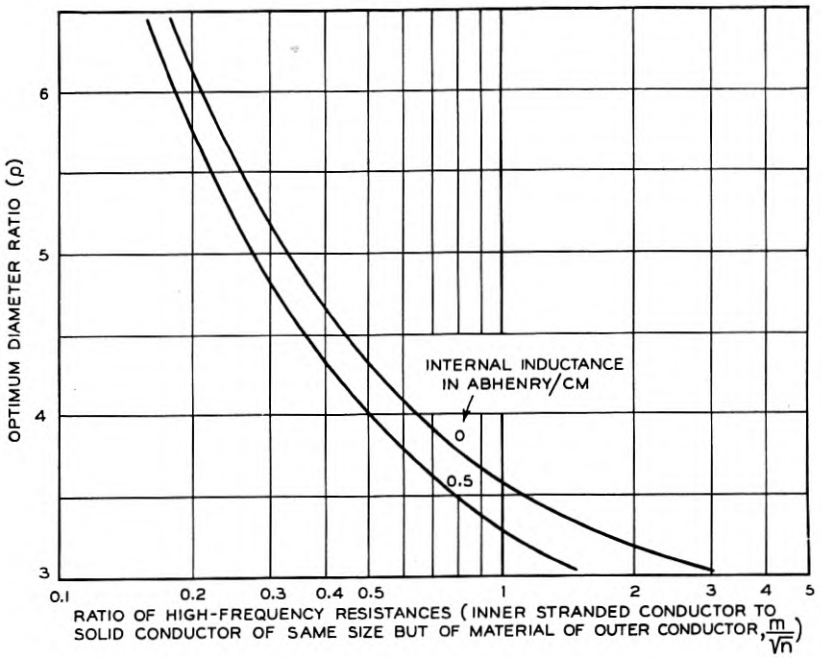


Fig. 6—Optimum diameter ratio of coaxial circuit with stranded inner conductor.

outer radius and that of the outer conductor as its inner radius, thus

$$R_i = \frac{k_1}{b} \qquad \text{and} \qquad R_0 = \frac{k_2}{c}. \qquad (27), (28)$$

These relations are approximately true for all the types of circuits which have been discussed. Let

$$r = \frac{k_2}{k_1} = \frac{R_0}{R_i} \cdot \frac{c}{b} = \frac{R_0}{R_i} \rho. \qquad (29)$$

If the internal inductance of the conductors is assumed to be zero, which is the most usual case, the high-frequency attenuation of the circuit may then be written

$$\alpha = \frac{k_1}{2c}(\rho + r)\frac{1}{2\log_e \rho}. \tag{30}$$

Upon minimizing with respect to c/b the condition for minimum high-frequency attenuation is found to be

$$\log_e \rho = \frac{\rho + r}{\rho} = 1 + \frac{R_0}{R_i}. \tag{31}$$

These relations have been found useful in certain instances.

Balanced Shielded Circuits

Though arrangements of three or more coaxial conductors are possible,[14] practical interest is almost wholly limited to coaxial circuits employing but two conductors. With balanced shielded circuits, however, the number of conductors, counting the shield as one, is necessarily three and may be more. With a coaxial circuit, moreover, the cylindrical shape is the natural and usual one for the conductors. With balanced shielded circuits, on the other hand, there enter a number of possibilities. Not only are cylindrical shapes of conductors and shield to be considered, but a variety of other shapes as well. More complex, therefore, than the foregoing problems in optimum proportioning are those for balanced shielded circuits, now to be discussed.

Shielded Pair—Cylindrical Conductors and Shield

The simplest form of balanced shielded circuit is a shielded pair comprising two cylindrical conductors surrounded by a cylindrical shield. Such a circuit is shown diagrammatically in cross-section in Fig. 7. For the present, attention will be directed to the circuit obtained when the two enclosed conductors are connected one as a return for the other.

Condition for Minimum Attenuation [15]

As before, it is desired to minimize the high-frequency attenuation. Let it be assumed first, as in the coaxial circuit, that the area within the shield is fixed, the conductors are of solid material and the dielectric is gaseous. Let b represent the radius of each conductor in Fig. 7, c the inner radius of the shield, h the distance from the center of either conductor to the center of the shield, λ_1 the conductivity of each con-

ductor, λ_2 that of the shield, and n the ratio of λ_1/λ_2. Expressions for the high-frequency attenuation of this circuit have been given in unpublished formulas developed by S. A. Schelkunoff and by Mrs. S. P. Mead. The approximate formula given below is due to the latter.

$$\alpha = \frac{\rho\left[1 + \frac{1+2\nu^2}{4\nu^4}(1-4\sigma^2)\right] + 4\sqrt{n}\,\sigma^2\left[1+\sigma^4 - \frac{1+4\nu^2}{8\nu^4}\right]}{\log_e\left[2\nu\frac{1-\sigma^2}{1+\sigma^2}\right] - \frac{1+4\nu^2}{16\nu^4}(1-4\sigma^2)} \times \frac{1}{4c}\frac{\sqrt{f\epsilon}}{\sqrt{\lambda_1}} \text{ nepers per cm.,} \quad (32)$$

where $\sigma = h/c$ and $\nu = h/b$.

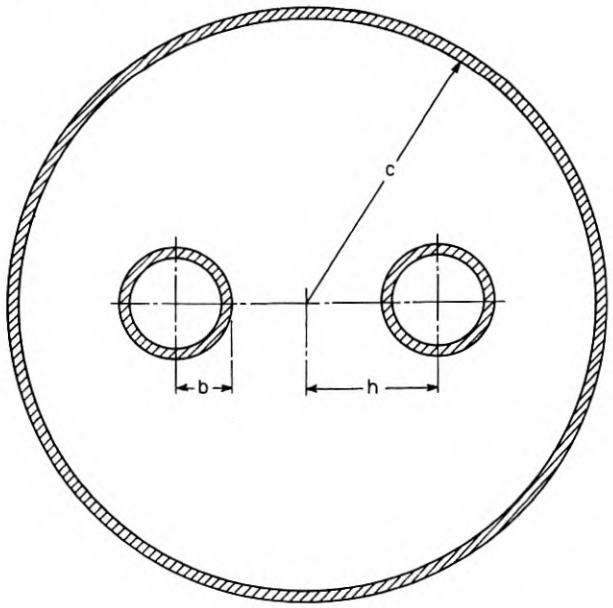


Fig. 7—Shielded pair.

The values of the diameter ratio (ρ) and what may be termed the spacing ratio (σ), which make this expression a minimum for different values of the conductivity ratio n, can be determined in different ways. One possible method is to find the values of h and b which satisfy the equations $\partial\alpha/\partial h = 0$ and $\partial\alpha/\partial b = 0$. The partial derivatives are, however, very complicated. Accordingly a preferable alternative is to substitute various pairs of values of ρ and σ in (32) and determine, graphically or otherwise, the particular pair which makes it a minimum. In this way it is found that when the conductors and shield are of the

same material, so that n equals unity, the optimum values are approximately

$$\rho = \frac{c}{b} = 5.4; \qquad \sigma = \frac{h}{c} = .46. \qquad (33), (34)$$

The optimum diameter and spacing ratios for different values of the conductivity ratio n are shown in Figs. 8 and 9. For copper conduc-

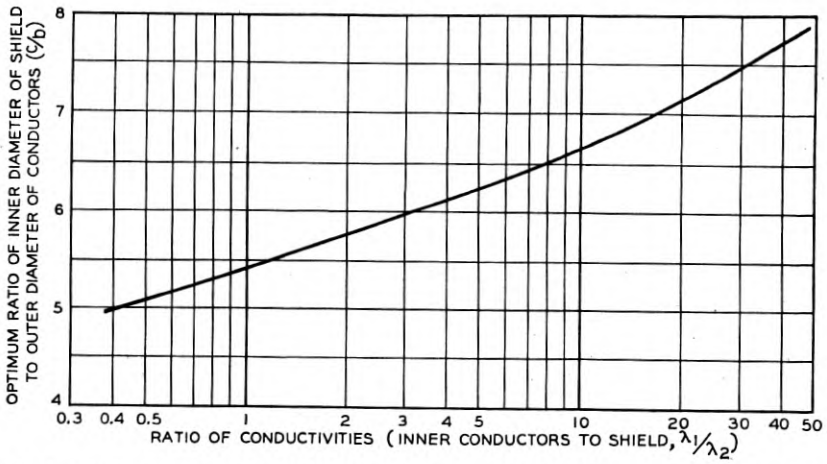


Fig. 8—Variation of optimum diameter ratio of shielded pair with conductivity ratio.

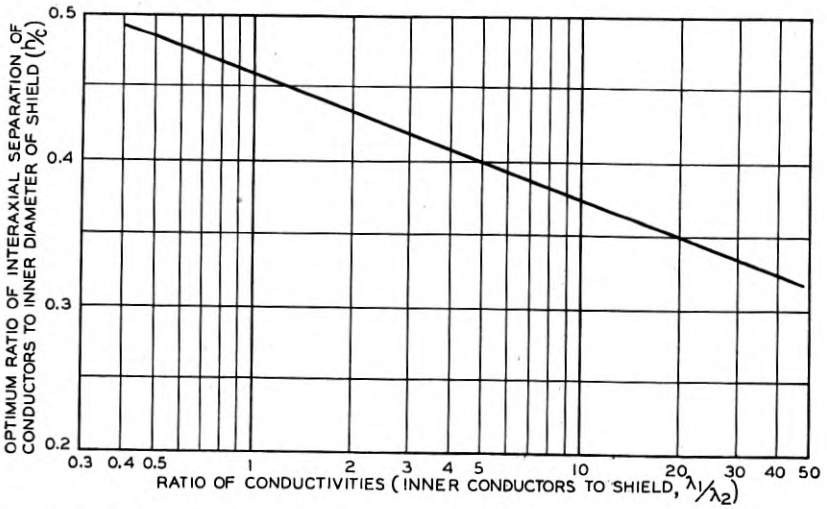


Fig. 9—Variation of optimum spacing ratio of shielded pair with conductivity ratio.

tors and a lead shield, the values are approximately 6.9 and .36, respectively.

As with the coaxial circuit, these optimum relations are independent of the diameter and thickness of the shield. Hence they make it possible to find the minimum size of shield necessary for a given value of high-frequency attenuation. The optimum relations are also independent of the frequency, provided the frequency is high enough for the approximate formulas to hold. The inner conductors may be either hollow or solid.

Condition for Maximum Characteristic Impedance[15]

Occasionally it is of interest to know the condition that must be satisfied to obtain maximum high-frequency characteristic impedance for a solid pair with circular shield. At high frequencies the value of $1/\sqrt{LC}$ approaches a constant value equal to the velocity of light divided by the square root of the ratio of the dielectric constant of the circuit to that of air. Hence the condition for maximum characteristic impedance is also, from equation (11), that for maximum inductance and minimum capacitance.

Accordingly, the high-frequency characteristic impedance of the shielded solid pair circuit is given by the formula:

$$Z_0 = \frac{4}{\sqrt{\epsilon}}\left(\log_e \left[2\nu \frac{1-\sigma^2}{1+\sigma^2} \right] - \frac{1+4\nu^2}{16\nu^4}(1-4\sigma^2) \right) \text{ abohms.} \quad (35)$$

Let it be assumed first that the wires are very small compared with the shield. Then equation (35) may be written

$$Z_0 = \frac{4}{\sqrt{\epsilon}} \log_e \left[\sigma \frac{1-\sigma^2}{1+\sigma^2} \right] + \frac{4}{\sqrt{\epsilon}} \log_e 2\rho \text{ abohms.} \quad (36)$$

For a given ratio of inner diameter of shield to outer diameter of conductor, the second term of this expression is constant. By minimizing the first term with respect to σ, it is found that, so long as the ratio of inner diameter of shield to conductor diameter is large, maximum characteristic impedance is obtained when σ has a value of .486.

If the conductors are large compared with the shield, equation (36) no longer holds. However, since the capacitance and high-frequency characteristic impedance are inversely proportional to one another, the position of the conductors with respect to the shield must be such as to minimize the capacitance. It is clear that as the conductor diameter approaches the inner radius of the shield, σ approaches 0.5 for minimum capacitance. Hence, for any ratio of inner diameter of shield to

diameter of conductor, the ratio of the interaxial separation of the conductors to the inner diameter of the shield which gives maximum characteristic impedance lies between the limits 0.486 and 0.500. For practical purposes a value of about 0.49 may generally be used.

Effect of Dielectric

The effect of dielectric for a shielded pair is similar to that for a coaxial circuit. When the insulation is so disposed between conductors and shield that a line of dielectric flux passes through only one kind of dielectric material, the second term of the attenuation formula is independent of the proportioning of conductors and shield, so that the optimum proportions as given in Figs. 8 and 9 are unchanged. These values will also serve for most practical cases where a line of dielectric flux may pass through more than one kind of material.

Effect of Frequency

At frequencies where the approximate formulas no longer hold, the conditions for minimum attenuation as given by Figs. 8 and 9 undergo some change, especially the former. As the frequency is decreased the attenuation is minimized by increasing the size of conductor for a given size of shield. In other words, the optimum diameter ratio grows less. The optimum spacing ratio increases from 0.46 toward the value which gives minimum capacitance, i.e., approximately 0.49.

Pair in Space

It is interesting to digress for a moment to consider briefly the case shown in Fig. 10 of a pair of round conductors in space. This may

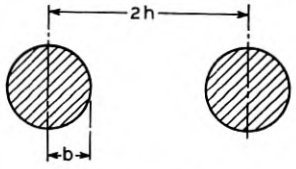


Fig. 10—Pair in space.

be regarded as a pair surrounded by a shield of infinite diameter. If the conductors are of solid material, the attenuation of the circuit at high frequencies is

$$\alpha = \frac{P}{4b}\sqrt{\frac{f\epsilon}{\lambda}}\,\frac{1}{\cosh^{-1}\nu}\ \text{nepers per cm.}, \qquad (37)$$

where P is the proximity effect factor, given in a paper by J. R. Carson.[16] At high frequencies this factor reduces to the asymptotic value

$$P = \frac{\nu}{\sqrt{\nu^2 - 1}}. \tag{38}$$

For a given high frequency and given wire separation, assuming the dielectric constant and conductivity to be fixed, equation (37) becomes

$$\alpha = \frac{K_3 \nu^2}{\sqrt{\nu^2 - 1}\cosh^{-1}\nu}, \tag{39}$$

where K_3 is a constant.

For a given wire separation this expression is minimized when

$$\nu = \frac{h}{b} = 2.27. \tag{49}$$

For open-wire pairs, which may be considered as approaching pairs in space, it is ordinarily cheaper to obtain any desired attenuation at a given frequency by using a wide separation and relatively small conductors rather than a narrow separation and conductors of such size as to satisfy (40). This relation is of considerable utility, however, in that it is a reasonably close approximation to the optimum for many kinds of shielded pairs. The corresponding ratio for the shielded solid pair, as given by (33) and (34), is approximately 2.5.

Shielded Stranded Pair

The preceding discussion of shielded pairs has been limited to types of enclosed conductors such that high-frequency currents are crowded toward the conductor surfaces. There will now be found the optimum proportioning when the enclosed conductors are stranded.[17]

The capacitance and inductance between two shielded stranded wires when surrounded by a cylindrical shield are approximately

$$C = \frac{\epsilon}{4\log_e\left[2\nu\frac{1-\sigma^2}{1+\sigma^2}\right]} \text{ abfarads per cm.}, \tag{41}$$

$$L = 4\log_e\left[2\nu\frac{1-\sigma^2}{1+\sigma^2}\right] + 2L_i \text{ abhenries per cm.}, \tag{42}$$

where L_i is the internal inductance of each conductor.

If it be assumed that the current distribution is uniform over the cross-section of the enclosed conductors, the resistance of each is the

same as if its return were coaxial. Hence the high-frequency resistance of each conductor is

$$R_i = \frac{2m}{b}\sqrt{\frac{f}{\lambda_1}} \text{ abohms per cm.} \tag{43}$$

The high-frequency resistance of the shield can be shown to be

$$R_0 = \frac{8ch^2}{c^4 - h^4}\sqrt{\frac{f}{\lambda_2}} \text{ abohms per cm.} \tag{44}$$

The high-frequency attenuation of the shielded stranded pair, found by substituting equations (41) to (44) in (1), is, with zero dielectric loss,

$$\alpha = \frac{\frac{m}{2c}\sqrt{\frac{f}{\lambda_1}}\left[\rho + \frac{4\sqrt{n}\sigma^2}{m(1-\sigma^4)}\right]}{\sqrt{\left[\log_e 2\nu\frac{1-\sigma^2}{1+\sigma^2}\right]\left[4\log_e 2\nu\frac{1-\sigma^2}{1+\sigma^2} + 2L_i\right]}} \text{ nepers per cm.} \tag{45}$$

The optimum proportions of the shielded stranded pair at high frequencies depend, therefore, on the two quantities $m/\sqrt{n}$ and L_i. For any given shield radius c, the values of h and b which give minimum attenuation may be found by setting

$$\frac{\partial\alpha}{\partial h} = 0; \quad \text{and} \quad \frac{\partial\alpha}{\partial b} = 0. \tag{46), (47}$$

By imposing the first condition it is found that

$$\frac{\partial M}{\partial h}\frac{(c^4 - h^4)^2}{8ch(c^4h^4)} = M\frac{2\log_e M(4\log_e M + 2L_i)}{8\log_e M + 2L_i}\frac{1}{\frac{m}{\sqrt{nb}} + \frac{4ch^2}{c^4 - h^4}}. \tag{48}$$

Imposing the second condition we find that

$$-\frac{\sqrt{nb^2}}{m}\frac{M}{b} = M\frac{2\log_e M(4\log_e M + 2L_i)}{8\log_e M + 2L_i}\frac{1}{\frac{m}{\sqrt{nb}} + \frac{4ch^2}{c^4 - h^4}}, \tag{49}$$

where $M = 2\nu(1 - \sigma^2)/(1 + \sigma^2)$.

Upon equating the left hand members of (48) and (49), and substituting the values of the derivatives, the following expression results

$$\rho = \frac{8\sigma^2(1 + \sigma^4)}{\frac{m}{\sqrt{n}}(1 - \sigma^4)(1 - 4\sigma^2 - \sigma^4)}. \tag{50}$$

This expression is the locus of values of the ratio ρ which give minimum attenuation for different assumed values of the ratio σ. The unique values of $h/c = \sigma$ and $c/b = \rho$, which give minimum attenuation for a given value of $m/\sqrt{n}$ and L_i, may be obtained by taking pairs of σ and ρ which satisfy equation (50), substituting them in equation (45), and graphically determining the pair for which the attenuation is a minimum.

Figures 11 and 12 show a graph, obtained in this way, of the optimum

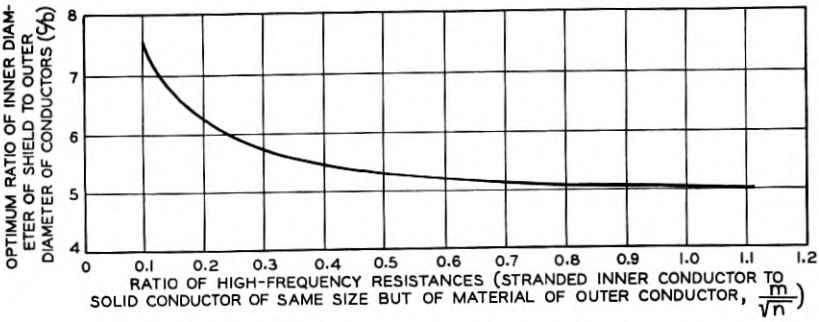


Fig. 11—Optimum diameter ratio of shielded stranded pair.

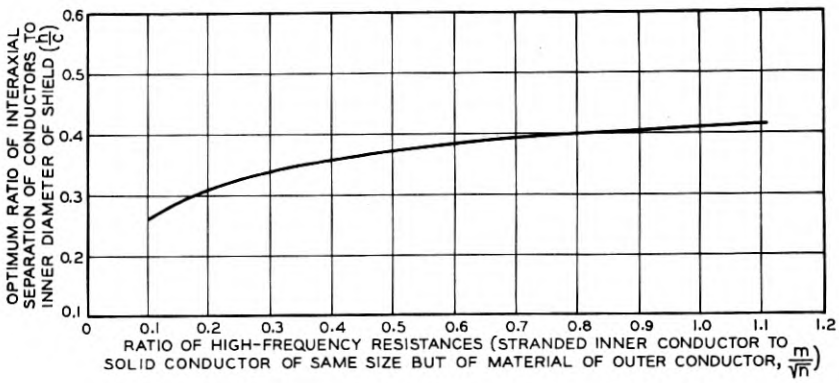


Fig. 12—Optimum spacing ratio of shielded stranded pair.

proportions for a shielded stranded pair, plotted as a function of $m/\sqrt{n}$ for a value of L_i equal to 0.5 abhenry per centimeter, which corresponds to the case where each conductor is completely stranded.

Pair with Shield Return

The discussion of the shielded pair thus far has been concerned solely with the circuit which employs one of the enclosed conductors as a return for the other. A second circuit may be obtained by

transmitting over the two enclosed conductors in parallel with the shield as the return. This latter circuit alone is less efficient than a coaxial circuit formed by replacing the two inner conductors with a single one. If, however, the two circuits obtainable from the shielded pair structure can both be employed without excessive mutual interference, there will be a considerable increase in the usefulness of the system, measured in terms of the total frequency range that can be transmitted without exceeding a given attenuation. It is therefore of interest to determine the conditions making the total transmitted frequency range for the two circuits a maximum.[18]

The high-frequency attenuation of each circuit, assuming solid conductors, can be written

$$\alpha = K\sqrt{f}, \tag{51}$$

where K is a constant, different for each circuit, which depends on the size and material of the conductors, and the dielectric constant of the insulation. Leakage is assumed negligibly small.

Using subscripts 1 and 2, respectively, to designate the circuit comprising the two enclosed conductors one as a return for the other and the circuit comprising the two wires in parallel with shield return, it follows that

$$f_1 + f_2 = \frac{\alpha_1^2}{K_1^2} + \frac{\alpha_2^2}{K_2^2}. \tag{52}$$

Letting $A = \alpha_2/\alpha_1$

$$f_1 + f_2 = \alpha_1^2 \left(\frac{1}{K_1^2} + \frac{A^2}{K_2^2} \right). \tag{53}$$

Equation (53) gives the sum of the frequency ranges that can be transmitted in the above manner over any given shielded pair for any given attenuation at the highest frequencies of the bands. To obtain maximum total range, this equation must be maximized.

The attenuation of the circuit comprising one enclosed conductor as a return for the other is given by equation (32), from which the value of K_1 can be obtained immediately. An expression for K_2 has been given in an unpublished formula due to Mrs. S. P. Mead, as follows:

$$K_2 = \frac{U + V}{\log_e \left[\frac{\rho(1 - \sigma^4)}{2\sigma} \right] - \frac{1 + 4\sigma^4}{1 + 4\nu^2} \left(1 + 4\sigma^4 \left(\frac{5 + 4\nu^2}{1 + 4\nu^2} \right) \right)} \frac{1}{4c} \sqrt{\frac{\epsilon}{\lambda_1}} \tag{54}$$

in which

$$U = \rho \left[1 + \frac{8\nu^2(1 + 4\sigma^4)}{(1 + 4\nu^2)^2} \left(1 + 4\sigma^4 \left(\frac{9 + 4\nu^2}{1 + 4\nu^2} \right) \right) \right],$$

$$V = 2 \left[1 + 2\sigma^4 + \frac{8\sigma^4}{1 + 4\nu^2} \left(1 + 8\sigma^4 \left(\frac{5 + 4\nu^2}{1 + 4\nu^2} \right) \right) \right].$$

For a given inner radius of shield and given dielectric constant and conductor material, the diameter and spacing ratios which make equation (53) a maximum can be obtained by the substitution method previously described. When the conductors and shield are of the same material and $A = 1$, computation shows that the total frequency range is a maximum when the radius ratio ρ equals 5.9 and the spacing ratio σ equals 0.33. This value of $A = 1$ represents an important practical case, since it will, as a rule, be desirable to employ the same repeater points for each circuit and permit the same attenuations between repeater points. It is also of interest, however, to determine the effect of other values of A.

When A is zero, the problem reduces to that of the simple shielded pair, which has been shown previously to be minimized by the proportions given in (33) and (34).

When A becomes large, or, in other words, when the phantom circuit alone is used, $1/K_2^2$ must be maximized. It is obviously necessary that the enclosed conductor be in contact and, accordingly, the spacing ratio must be the reciprocal of the diameter ratio. For this condition the following proportions result:

$$\rho = \frac{c}{b} = 6.0; \qquad \sigma = \frac{h}{c} = 0.17. \qquad (55), (56)$$

The above proportions are optimum only when the enclosed conductors and the shield are of the same conductivity. The relations for the case of unequal conductivities may be derived in a similar manner. For practical purposes the effect of dielectric loss on the optimum proportions is negligible.

Double Coaxial Circuit

Another form of balanced and shielded transmission circuits may be obtained by using two coaxial conductor units, the transmission path consisting of the two inner coaxial conductors in series, with the outer coaxial conductors serving only for shielding. Such a circuit is shown diagrammatically in cross-section in Fig. 13. Usually the outer conductors would be in practically continuous contact with each other. A circuit of this type will handle a frequency band extending to lower values than can be used with a single coaxial circuit, since it is balanced and the two coaxial units can be transposed by twisting or by periodic interchange of their positions. At high frequencies, where the shielding of the outer conductor of the coaxial circuit becomes effective, the outer conductors may be separated to any desired distance. It is essential, however, that they be connected together at the ends of the circuit.

In such a double coaxial circuit, used at high frequencies, equal and opposite currents will flow on the inner and outer conductors of each coaxial unit. The resistance and inductance of the balanced circuit will, therefore, be twice, and the capacitance and leakance one-half, the corresponding values for one coaxial unit. The attenuation of the balanced circuit is equal to the attenuation of one coaxial unit and may be expressed by the formulas previously given, where the various symbols are understood to refer to one unit of the circuit. Accordingly the optimum high-frequency proportions are the same as those previously derived for ordinary coaxial circuits of different types.[19]

As the frequency is reduced, the optimum proportions become different from those for coaxial units, since the circuit inductance ap-

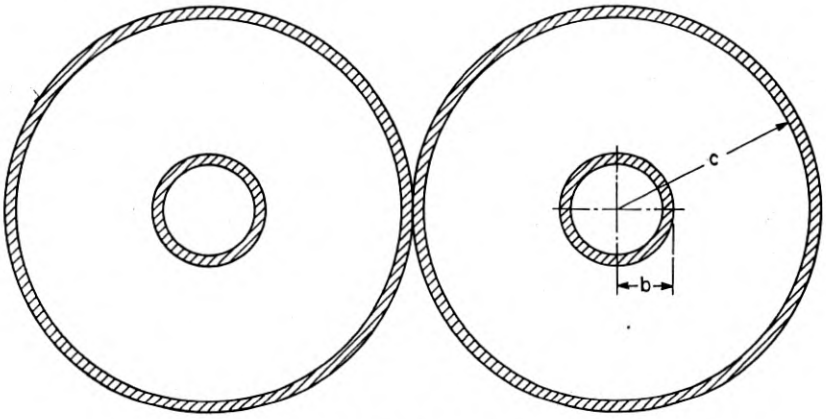


Fig. 13—Double coaxial circuit.

proximates more closely that for a simple pair of wires occupying the positions of the inner conductors, while the capacitance remains equal to one-half of that of one coaxial unit. As a result the optimum diameter ratio is larger than at high frequencies.

Shielded Pair—Round Conductors and Oval Shield

The shield around a pair does not have to be cylindrical. Upon consideration of a pair of round conductors with a cylindrical shield, as shown in Fig. 7, it is evident that the shield approaches quite close to the conductors at the sides, while it is well removed from them at the top and bottom of the figure. This means that for a given area enclosed by the shield the capacitance of the circuit is greater than would be the case if the shield were kept at a more nearly uniform

distance from the conductors. Consequently, for a given area circumscribed by the shield, a reduction of attenuation can be secured by changing the shape of the shield.

The problem of determining the shape of shield which gives minimum high-frequency attenuation presents extreme difficulty, and a rigorous solution has not been obtained. However, it appears that a close approach to the ideal shape can be obtained by a shield having the cross-section shown in Fig. 14, which consists of two semi-circles

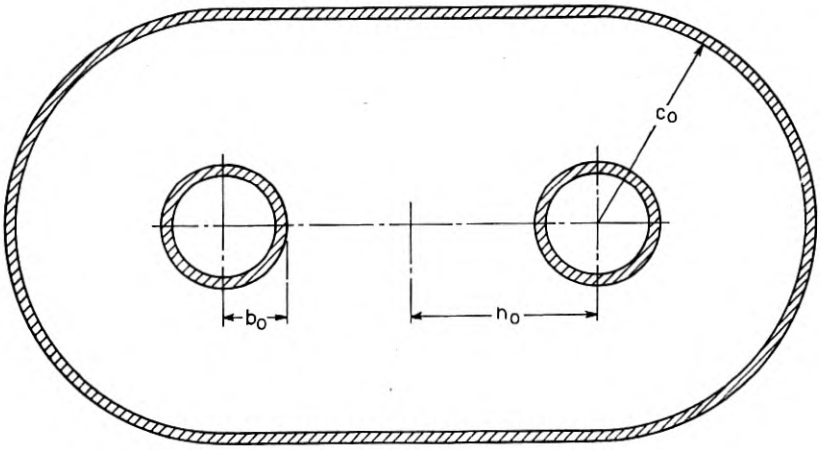


Fig. 14—Oval shielded pair.

joined by straight lines, the inner conductors being placed at the centers of the semi-circles. For convenience this shape of shield will be termed "oval."

The optimum proportioning [20] for a pair of conductors with such an oval shield may be closely approximated by comparison with the pair with circular shield and with the double coaxial circuit. In such comparison the cross-sectional areas of the different circuits will be assumed equal.

Consideration will first be given to the case where the enclosed conductors in Fig. 14 are of solid material. The conductivity of the conductors will be assumed the same as that of the shield, it being apparent that the same methods may be employed in the case of different conductivities. In arriving at the spacing ratio of the conductors for minimum attenuation, the condition for minimum capacitance will be used as a stepping stone. The spacing ratio of the conductors in Fig. 14 may be represented by $h_0/(c_0 + h_0)$. Comparison with Fig. 7 shows that the corresponding ratio for that figure is h/c, which,

it has already been seen, should have a value of approximately .486 for minimum capacitance. It is evident that the value of the ratio $h_0/(c_0 + h_0)$ for minimum capacitance in Fig. 14 should be very close to .486, but in view of the concentricity of the conductors with the semi-circular parts of the shield it should be slightly less than this value. It has been found that to obtain minimum capacitance for an oval shielded circuit the spacing ratio should be approximately

$$\frac{h_0}{c_0 + h_0} \doteqdot .47. \tag{57}$$

It has been seen for Fig. 7 that to obtain minimum high frequency attenuation the spacing ratio is shifted from the value of .486, which gives minimum capacitance, to a value of about .46. For Fig. 14, however, the current density in the shield is more uniform, so that the proximity effect between conductors is less completely compensated by the shield currents. Hence the spacing ratio for minimum high-frequency attenuation for the oval shielded circuit should be approximately the same as that for minimum capacitance, as given in (57) above.

There remains to be determined the second condition for minimum high-frequency attenuation for an oval shielded circuit of given cross-sectional area, namely, the optimum value of the diameter ratio c_0/b_0. Comparison with Fig. 13 indicates that the optimum value of this ratio should be fairly close to the optimum value of 3.6 for the coaxial circuit. Comparison with Fig. 7, c_0 being equal to about $.69c$ for equal areas in the two cases, shows that the optimum value of the ratio c_0/b_0 should be slightly greater than 3.6. For practical purposes the optimum may be taken as

$$\frac{c_0}{b_0} \doteqdot 3.7. \tag{58}$$

With this ratio the size of the conductors with oval shield is, for the same cross-sectional area, approximately the same as that of the optimum size of conductors with circular shield.

The capacitance of the pair with oval shield is smaller than the capacitance of the pair with circular shield, because the inner conductors of the former are more widely separated and are farther from the shield. It is very slightly larger than the capacitance of the double coaxial circuit.

The part of the resistance of the oval shielded circuit which is due to the shield will be less than that for a circular shield because of the more uniform current density in the shield. However, as has been

noted, the proximity effect between conductors is less completely neutralized by the shield currents than is the case for the circular shield. It appears that these two effects may approximately balance one another, and that the circuit resistance is approximately the same for both oval and circular shielded circuits.

It is found that a circuit of approximately optimum proportions comprising two solid round wires surrounded by an oval shield has about 12 per cent lower attenuation than a circuit with circular shield of equal cross-sectional area.

When the conductors enclosed within the oval shield are stranded there is no increase of conductor resistance due to proximity effect. On this account it is desirable to bring the conductors closer together in order to reduce the shield loss and the optimum spacing ratio will be less than for the case of solid conductors. With stranded conductors the attenuation reduction as compared with the circular shield is greater than in the case of solid wires; for example, if the resistance ratio (m) is .7, the attenuation with oval shield will be about 25 per cent less than that of the circular shield.

The circular form of shield is ordinarily the most convenient and practical one. A disadvantage of an oval shield as compared thereto is unequal stiffness or resistance to bending in different directions.

Shielded Pair—Quasi-Elliptical Conductors

It has been suggested at different times that the ordinary round form of conductor, while well adapted for manufacturing purposes, may not be the theoretically optimum shape for many types of high-frequency transmission circuits. Speculations in this respect have differed greatly, and a large variety of non-circular shapes of conductors have been proposed, including flat strips, strips with concave or convex faces opposite one another, angular forms, etc. However, except in the case of the coaxial circuit, for which the circular form is clearly the optimum, there has been, so far as the authors are aware, no exact analytical determination of the optimum conductor shape for a given type of circuit.

A complete treatment of possible problems of this kind would extend to great length. It is worth while, however, to consider a single problem, namely, that of determining what shape and spacing for a pair of conductors with circular shield will result in minimum high-frequency attenuation. This problem is of particular interest inasmuch as the circular shape is ordinarily the most convenient and practical one for a shield.

In attacking this problem the fundamental principles which deter-

mine the high-frequency attenuation of a circuit comprising a pair of conductors surrounded by a shield may be briefly examined. At high frequencies, where the currents are crowded toward the surfaces of the conductors, the attenuation is proportional to the product of the resistance and capacitance of the circuit, both of which are functions of the flux density in the dielectric.

With a circular shield, and round conductors, the flux density is far from uniform around the surfaces of the conductors, being relatively high at points nearest the shield and also at points nearest the shield's center, and a minimum at points about half-way between. Accordingly, it appears that the high-frequency resistance of the conductors can be reduced by reshaping them so as to make the flux distribution more uniform. This can be accomplished by squeezing the conductors at regions of maximum flux density and bulging them at regions of minimum flux density, thereby producing a conductor of approximately elliptical cross-section.

The flux distribution around the shield is also far from uniform, being a maximum at points nearest the conductors and a minimum at points 90 degrees away. Making the enclosed conductors elliptical tends to reduce this non-uniformity, thereby reducing the circuit resistance due to loss in the shield.

This process of reshaping the conductors can not be carried very far, however, because it soon increases the circuit capacitance more than it decreases the resistance. It is difficult to treat this problem by rigorous mathematics, but an analysis can be made which yields an approximate solution.

For certain conductor shapes, the high-frequency attenuation of a pair with circular shield may be determined by a method involving the substitution of charged filaments for the conductors. Let any number of positively and negatively charged filaments be included in the shield, the net charge on the filaments being zero. The electrostatic potential at any point of this system can readily be determined by known methods. Thus, for example, Fig. 15 shows the location of the equipotential surfaces for the case of two oppositely charged filaments placed within a circular shield, the distance from each filament to the center of the shield being .46 times the shield radius.

In any such system, a conducting cylinder whose external surface corresponds to, and whose potential is equal to the potential of, a particular equipotential surface may be substituted for the part of the system contained within that surface without disturbing the flux distribution external to it. Consequently, the capacitance of a shielded circuit employing equal and oppositely charged conductors

having the same shape as any two corresponding equipotential surfaces of the electrostatic system can be determined.

The flux density at any points on the conductors or on the shield is proportional to the rate of change of the potential with respect to the normal to the surface at that point. The high-frequency resistances of the conductors and shield, respectively, are proportional to the

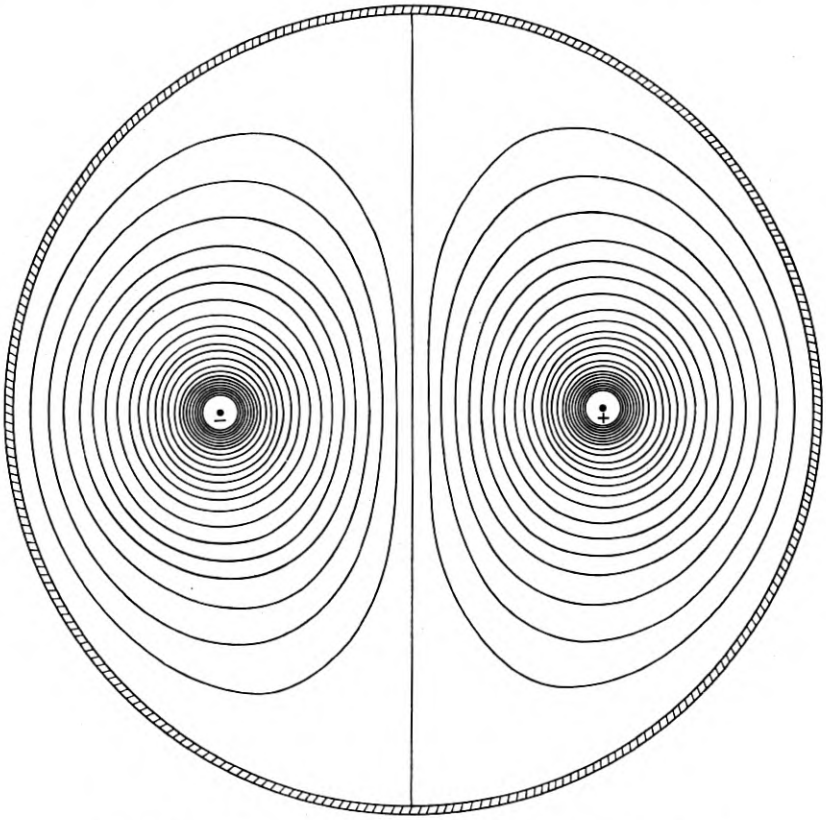


Fig. 15—Equipotential lines around shielded charged filaments.

integral of the square of the flux density around their periphery. Thus the high-frequency resistance of the circuit may be determined, and from this and the capacitance, the high-frequency attenuation.

This method makes it possible to determine and compare the high-frequency attenuations of conductors having shapes corresponding to the equipotential surfaces for various assumed arrangements and numbers of charged filaments. If, however, the problem be that of

determining the attenuation for a given shape of conductor, there may be great difficulty in finding the arrangement and number of filaments which will produce an equipotential surface to coincide with the given shape.

By applying this method to a series of approximately elliptical conductors previously shown to be of the shape that would be expected to have lower attenuation than circular conductors, what is considered a close approximation to the optimum shape of conductor for a pair with circular shield has been arrived at. This is approximately an ellipse whose major axis is about 5 per cent longer than its minor axis, the latter being in line with the center of the shield. The high-frequency attenuation of a circuit with circular shield and conductors of this shape is approximately 2 per cent lower than that for the same shield with round conductors. This reduction does not appear enough to offset the practical difficulties involved with conductors of such shape.

Shielded Quad

The number of conductors enclosed within a shield, instead of being one, as in the coaxial, or two, as in the shielded pair, may be more. By placing four conductors within a common shield, two separate balanced-to-ground circuits may be obtained. If sufficiently good balance can be obtained between these circuits, the total frequency band which can be transmitted within a given cross-sectional area may be increased. To obtain balance, the plane of the conductors of one circuit needs to be at right angles to that of the other circuit and all conductors should be equidistant from the axis of the shield. The pairs may be twisted or spiralled about the axis of the shield.

An arrangement of this kind is shown in Fig. 16, where four round conductors are placed within a circular shield to form a shielded quad, or, as it is frequently described when the conductors are twisted, a "shielded spiral four." Diagonally opposite conductors are used as the sides of a circuit.

Approximate formulas for the high-frequency attenuation of either circuit of Fig. 16, when the enclosed conductors are solid, have been derived in unpublished work of Mrs. S. P. Mead and S. A. Schelkunoff. The optimum high-frequency proportioning of the system, assuming the same conductivity for both enclosed conductors and assuming gaseous dielectric, has been determined by Mrs. Mead. The results are shown in Figs. 17 and 18, where the optimum diameter ratio and spacing ratio are plotted as functions of the ratio of the conductivity of the enclosed conductors to that of the shield. For the case of

equal conductivities of conductors and shield the optimum values are

$$\rho = \frac{c}{b} = 6.8; \qquad \sigma = \frac{h}{c} = .49. \qquad (59), (60)$$

These values may be compared with 5.4 and .46, respectively, for the pair of round conductors with circular shield. The high-frequency attenuation of each shielded quad circuit with optimum design is,

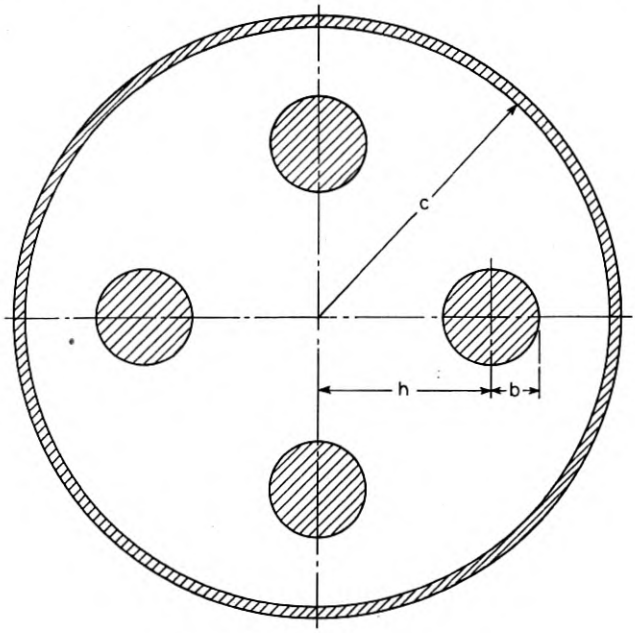


Fig. 16—Shielded quad.

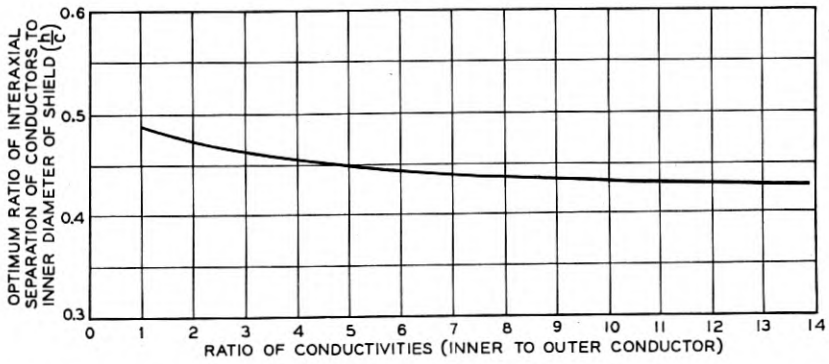


Fig. 17—Variation of optimum spacing ratio of shielded quad with conductivity ratio.

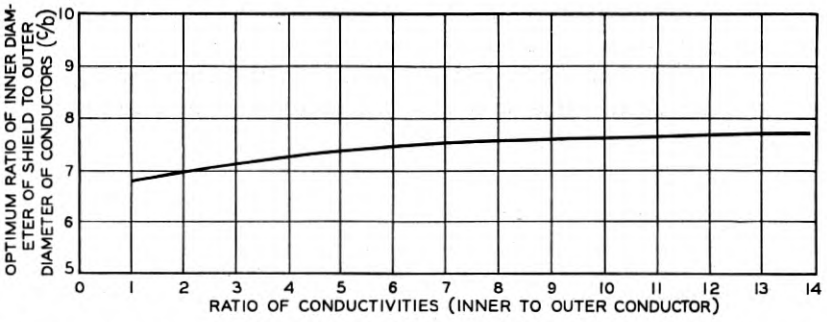


Fig. 18—Variation of optimum diameter ratio of shielded quad with conductivity ratio.

for the same diameter of shield, about 10 per cent higher than that of a shielded pair for its optimum design.

Conclusion

There have been discussed a number of different types of individually shielded circuits, both balanced and unbalanced, and the proportioning of these circuits for minimum high-frequency attenuation has been determined. The following table summarizes the optimum proportions for the more important circuits treated above. The values given are for the case where all the conductors are of the same material.

Circuit	Diameter Ratio (ρ)	Spacing Ratio (σ)
Simple coaxial	3.59	—
Double coaxial	3.59	—
Shielded pair, round conductors and circular shields	5.4	0.46
Shielded pair, round conductors and oval shield	3.7	0.47
Shielded quad	6.8	0.49

Of the transmission characteristics of these circuits, a property of particular interest is the attenuation, since, assuming adequate shielding, it is this which determines either the required repeater spacing for a given transmitted frequency band or the width of frequency band obtainable with a given repeater spacing. For each type of circuit considered there has been determined the ideal proportioning whereby the high-frequency attenuation of the circuit may be minimized. In addition a variety of methods for the solution of problems in optimum proportioning have been outlined.

It is, of course, feasible by adjustment of size to obtain the same high-frequency attenuation for all these different types of circuits. However, the size of a structure is usually reflected in its cost. An interesting picture can therefore be drawn by comparing the attenua-

tions, at the same high frequency, of different types of circuits having the same cross-sectional area and of the same material. For structures with solid wall conductors and air insulation the comparison works out as shown in the table below, the attenuation of the coaxial circuit being used as a standard of reference.

Coaxial circuit	1.00
Shielded pair, round conductors and circular shield	1.50
Double coaxial circuit	2.00
Shielded pair, round conductors and oval shield, approximately	1.3
Shielded pair, circular shield with quasi-elliptical conductors, approximately	1.47

In each case the cross-sectional area is taken as that enclosed within the shield. This neglects any differences in the thickness of shield that may be required.

A specific comparison of considerable interest is that between an unbalanced coaxial circuit and a shielded pair, the latter being taken as representative of shielded balanced circuits. The table shows that, for the same attenuation, the cross-sectional area included within the shield is larger for the shielded pair than for the coaxial circuit. On the other hand, the use of balance in addition to shielding is advantageous in that it reduces the amount of shielding needed. The shielded pair makes possible the utilization of the entire frequency range, if desired, whereas with a coaxial circuit it is necessary to discard the lower frequencies where it is uneconomical to provide adequate shielding.

A thorough-going comparison of the relative advantages and fields of application of the various types of circuits which have been discussed would extend to great length. Clearly a large number of factors enter into the choice of the configuration of shielded high-frequency circuit to be used in any given instance. These factors include the width of frequency band to be transmitted, the degree of shielding required, the relative economy of manufacture of different structures, etc. While a complete exposition of these factors has not been attempted, the principles of optimum proportioning which have been discussed should be helpful in selecting the best configuration to meet given requirements, and the particular configuration chosen should be made to conform reasonably closely to the optimum.

References

1. L. Espenschied and M. E. Strieby, "Systems for Wide-Band Transmission Over Coxial Lines," *Elec. Engg.*, Vol. 53, October, 1934, p. 1371; also *Bell System Technical Journal*, Vol. 13, October, 1934, p. 654.
2. A. B. Clark, "Wide Band Transmission Over Balanced Circuits," *Elec. Engg.*, Vol. 54, January, 1935, pp. 27–30; also *Bell System Technical Journal*, Vol. 14, January, 1935, p. 1.
3. British Patent No. 284,005, C. S. Franklin, January 17, 1928.

4. E. J. Sterba and C. B. Feldman, "Transmission Lines for Short-Wave Radio Systems," *I. R. E. Proc.*, Vol. 20, July, 1932, p. 1163; also *Bell System Technical Journal*, Vol. 11, July, 1932, p. 411.
5. S. A. Schelkunoff, "The Electromagnetic Theory of Coaxial Transmission Lines and Cylindrical Shields," *Bell System Technical Journal*, Vol. 13, October, 1934, p. 532.
6. K. S. Johnson, "Transmission Circuits for Telephonic Communication," New York, 1925.
7. Alexander Russell, "The Effective Resistance and Inductance of a Concentric Main," *Phil. Mag.*, Vol. 17, Sixth Series, April, 1909, pp. 524–552.
8. Sir William Thomson, "On the Theory of the Electric Telegraph," *Roy. Soc. Lond. Proc.*, Vol. 7, May 3, 1855, pp. 382–399.
9. E. I. Green, U. S. Patent No. 2,029,420, February 4, 1936.
10. E. I. Green and F. A. Leibe, U. S. Patent No. 2,029,421, February 4, 1936.
11. H. A. Affel and E. I. Green, U. S. Patent No. 1,818,027, August 11, 1931.
12. S. Butterworth, "Eddy Current Losses in Cylindrical Conductors," *Royal Soc. London Phil. Trans.*, Vol. 222, May, 1922, pp. 57–100.
13. F. A. Leibe and T. C. Rogers, U. S. Patent No. 2,034,047, March 17, 1936.
14. E. I. Green, U. S. Patent No. 1,854,255, April 19, 1932.
15. E. I. Green, H. E. Curtis and S. P. Mead, U. S. Patent No. 2,034,032, March 17, 1936.
16. J. R. Carson, "Wave Propagation Over Parallel Wires: The Proximity Effect," *Phil. Mag.*, Vol. 41, April, 1921, p. 627.
17. E. I. Green and H. E. Curtis, U. S. Patent No. 2,034,033, March 17, 1936.
18. H. E. Curtis and S. P. Mead, U. S. Patent No. 2,034,026, March 17, 1936.
19. E. I. Green, U. S. Patent No. 2,034,035, March 17, 1936.
20. E. I. Green and F. A. Leibe, U. S. Patent 2,034,034, March 17, 1936.

Hyper-Frequency Wave Guides—General Considerations and Experimental Results *

By G. C. SOUTHWORTH

A peculiar form of electrical propagation is described below. It makes use of extremely high frequencies—even beyond those generally employed in radio. In some respects it resembles ordinary wire transmission but unlike the latter there are no return conductors, at least of the usual kind.

In this transmission, electromagnetic waves are sent through guides made up either of an insulator alone or of an insulator surrounded by a conductor. In a special case this insulator may be air. There are at least four different types of waves or electrical configurations that may be propagated. One of them is such that theory indicates its attenuation through a hollow conductor continuously decreases with increase of frequency. Although the paper deals largely with the nature of this transmission, some of the fundamental pieces of apparatus used in experimental work are described. They include generators, receivers and wave-meters.

Introduction

THIS paper describes a novel form of electrical propagation by means of which extremely high-frequency waves may be transmitted from one point to another, through specially constructed wave guides. The guide used for this purpose may take any one of several different forms. It may be a hollow copper pipe, which for the higher frequencies now available would be about 3 or 4 inches in diameter, or possibly a somewhat smaller conducting tube filled with some insulating material combining high dielectric constant and low loss, or it may conceivably be a rod or wire of dielectric material.[1]

The phenomena involved in this form of transmission are exceedingly interesting and at first sight paradoxical for in some cases transmission is effected through a single wire of insulating material surrounded by metal in place of a pair of metal wires surrounded by insulation. In others the wire is made entirely of insulating material. In still others electrical effects are observed only on the interior of hollow metal cavities instead of the exterior only as is ordinarily experienced. In all cases there is no return current path, at least of the kind that is commonly assumed in ordinary transmission.

The frequencies appropriate for this form of transmission begin at the higher of those generally known as ultra-high frequencies that is, 2000 mc. ($\lambda = 15$ cm.) and extend to an indefinite upper limit possibly

* To be presented at joint meeting of Amer. Phys. Soc. and I.R.E., Washington D. C., April 30, 1936.

[1] The mathematical theory of these phenomena is given in a companion paper by J. R. Carson, S. P. Mead and S. A. Schelkunoff, this issue of the *Bell System Technical Journal*.

set by the properties of available materials. These have for convenience been called hyper-frequencies. When these electromagnetic waves are propagated through either of the two forms of guide mentioned above that incorporates a metal sheath, there is little or no external field and consequently little or no interference from static or other extraneous noises.

As already mentioned there is no return conductor, at least of the kind with which we are generally familiar in ordinary wire or coaxial cable transmission. Corresponding to this difference in physical structure there are striking differences in the character of the waves propagated. On the other hand, when we compare this transmission with radio, where there might at first sight appear to be great similarity, we find little or no correspondence, for it turns out that as regards both the velocity of propagation and attenuation per unit length, radio and wave guides follow quite different laws.

In answer to the natural question as to what practical use there may be for transmission methods of this type the following considerations may be of interest: The size of structure that may be used as a guide is directly proportional to the wave-length. It happens that in structures that are at all convenient in size, the necessary frequencies correspond approximately to the highest range now being tried out in radio. If the size of structure is further reduced to make it more economical for use for long distance transmission, it is then necessary to use frequencies above this range. Thus far these can be produced and handled only with serious difficulty. Although it is possible to reduce the size of the guiding structure for a given frequency by the use of a suitable dielectric we are met with a conflicting difficulty of producing at reasonable cost the necessary medium that will incorporate high dielectric constant with sufficiently low losses. The situation then is that the art at these extreme frequencies is not yet at a point which permits a satisfactory evaluation of practical use. However, for short distance transmission or for use as antennas or projectors of radio waves or for selective elements analogous in nature to the tuning elements so commonly used in radio, there are not the same economic conditions limiting the size of structure. For such uses, then, structures of this type deserve serious consideration.

Theory indicates that one of the four types of waves (designated below as H_0) has progressively less attenuation as its frequency is increased. It happens, however, that this type requires for a given guide a higher range of frequencies than any others. This puts it, therefore, in a frequency range where the art is even less developed than for the other types of transmission and where it is even more

difficult to evaluate the economic and practical problems. This paper will, therefore, confine itself to a discussion of some of the fundamental properties of wave guides derived either from calculation or experiment. These properties include characteristic impedance, attenuation and velocity of propagation as well as frequency, selectivity and radiation.

Nature and Properties of Wave Guides

Analysis has shown that there are many kinds of waves that may be propagated through cylindrical guides. However, four of them are of unusual interest and are such as merit special consideration at this time. All four have been experimented with in our laboratory and their more important characteristics have been determined. This experimental work has been paralleled by a mathematical theory [1] to which it conforms most satisfactorily.

A good mental picture of the nature of the waves propagated through guides can probably best be had by abandoning the ordinary concept of current electricity flowing in a "go and return" circuit in favor of that of lines of electric and magnetic force. This latter concept has, of course, always been applicable even for low-frequency transmission over parallel wires or coaxial conductors but due to its complexity in pictorial representation it has usually been avoided. In the form of transmission with which we are now concerned, the field point of view is almost necessary.

Figure 1 is a pictorial representation based on this point of view of the four types of waves mentioned above as found in a guide surrounded by a metallic conductor. In these models the lines of electric force have been represented by solid lines and the lines of magnetic force have been shown by dotted lines. In the longitudinal sections, the small open circles represent lines of force directed toward the observer. The solid circles represent lines directed away from the observer. The designations E_0, E_1, H_0 and H_1 are convenient reminders of certain characteristics of these waves.

The first two waves have been designated as electric because there is a component of electric force in the direction of propagation. For similar reasons the latter have been known as magnetic waves. Such a designation is, of course, rather arbitrary and should not be construed to mean that either component resides alone. It is true here as in other forms of electromagnetic waves with which we are generally familiar, that both the electric and magnetic components are essential to the very existence of the wave and that they may conveniently be considered as different aspects of the same thing.

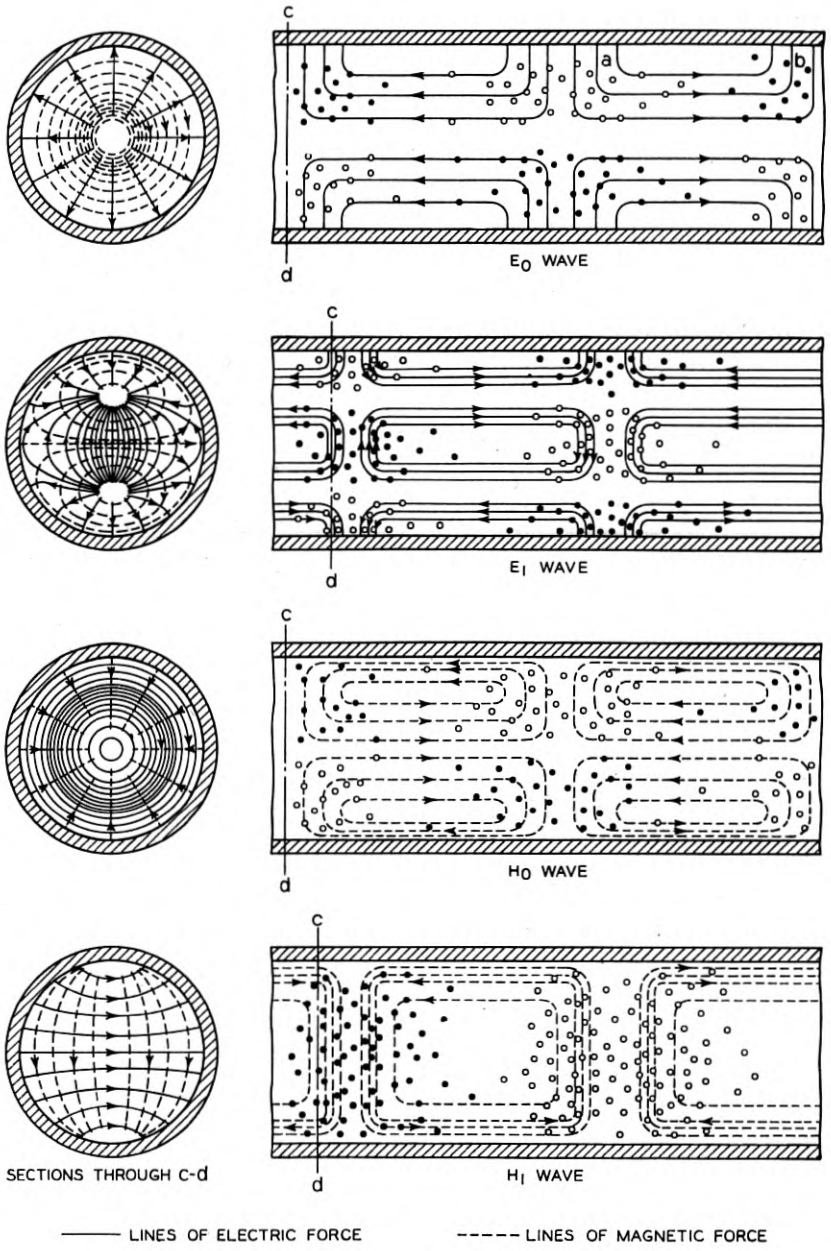


Fig. 1—Approximate configuration of lines of electric and magnetic force in a typical wave guide. Small solid circles represent lines of force directed away from observer. Propagation is assumed to be directed to the right and away from the observer.

Electromagnetic waves cannot be freely transmitted in dielectric wires or hollow conductors at all frequencies but only when the wave-length is less than a certain value set by the material of the guide and its dimensions. There is, therefore, for a given guide a critical frequency below which waves may not be propagated. We refer to this as the cut-off frequency. In a similar way we have for a given frequency, critical or cut-off diameters. These critical frequencies depend not only on the diameter (d) of the guide but on the dielectric constant (κ) of the medium as well. Also they are, in general, different for the different types of waves. For guides enclosed by a metallic conductor the cut-off wave-length is such that the circumference of the guide measured in wave-lengths is equal to the roots of certain Bessel's functions. These in turn result from solution of the Maxwell equations expressed in cylindrical coordinates. These relations are shown more fully in Table I.

TABLE I

Type of Wave	Bessel Functon	Root	Cut-off Wave-length $= \frac{\pi d\sqrt{\kappa}}{X}$
E_0	$J_0(X) = 0$	$X = 2.41$	$1.31d\sqrt{\kappa}$
E_1	$J_1(X) = 0$	$X = 3.83$	$0.82d\sqrt{\kappa}$
H_0	$J_0'(X) = 0$	$X = 3.83$	$0.82d\sqrt{\kappa}$
H_1	$J_1'(X) = 0$	$X = 1.84$	$1.71d\sqrt{\kappa}$

As a simple numerical example let us assume the H_1 type of wave having a frequency of 3000 mc. ($\lambda = 10$ cm.) being propagated in a hollow metallic pipe. The critical diameter turns out to be 5.85 cm. or roughly 2.30 inches. If the space were filled with a material having a dielectric constant of, say 5, this would have been reduced to a diameter of roughly one inch. For higher frequencies or for materials having still higher dielectric constants these critical dimensions would obviously be still further reduced and would be comparable in size to the larger conductors used in ordinary electrical practice. The critical dimensions for the other types of waves are of course larger.

Referring again to Fig. 1 we see that in the so-called E_0 type of wave, a line of electric force originating at a point a on the inner surface of the wall of the guide passes radially toward the center then axially and again radially to a corresponding point b on the inner wall roughly one-half wave-length farther along. The entire wave front as seen in cross section cut through cd consists of a symmetrical arrangement of these radial lines. The magnetic field associated with this wave consists of a series of coaxial circles shown as dotted lines not

unlike the magnetic field in a coaxial conductor such as shown in Fig. 2. The E_1 wave consists of electric and magnetic lines very similar in form to those associated with two parallel electric conductors surrounded by a metallic shield. The similarity between the fields for the two dielectric waves and the corresponding two arrangements for ordinary transmission is made more obvious by a comparison of Fig. 1 with Fig. 2. For the most part this similarity ends at this point, however, as their corresponding properties follow quite different laws.

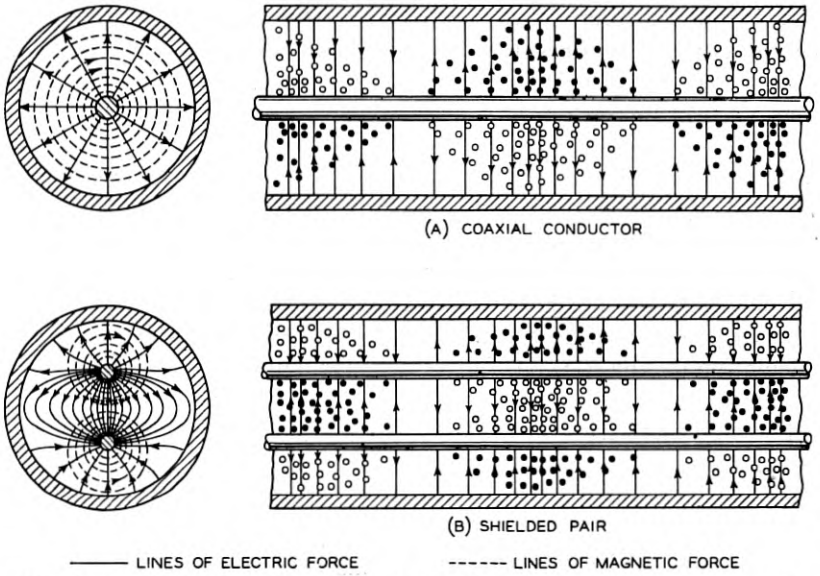


Fig. 2—Approximate configuration of lines of electric and magnetic force in a coaxial conductor and also in a shielded pair of conductors. Note similarity to E_0 and E_1 waves of Fig. 1.

The configurations of the two magnetic waves are somewhat similar to the electric waves provided we assume the electric and magnetic components to be interchanged. Nature has thus far failed to provide us with materials that possess exclusively magnetic conductivity in the sense that copper possesses electrical conductivity so there are no counterparts of Fig. 2 applicable to magnetic waves.

The general shape of the lines of electric force for all of these types of waves have been calculated. These fields have also been verified experimentally by means of a small probe consisting of a crystal detector with short pick-up wires connected to a sensitive meter. This probe was carried over the cross-section of the guide always

orienting the detector to obtain maximum deflection. These data confirmed not only the directions of the lines of force but their relative density as well.

There is one characteristic of the H_0 and H_1 configurations that at first sight seems inconsistent with our more usual views of electricity. It is the existence of a substantial tangential component of electric force apparently in close proximity to a metallic conductor. It must be borne in mind however that these frequencies are extremely high and that these distances after all represent an appreciable part of a wave-length.

If any of the four types of waves depicted above are propagated through a wire of dielectric material without the metal enclosure, lines of electric force which previously attached themselves to the inner walls of the sheath, in general, extend into the surrounding space and close as loops. This means that as the wave moves along the guide a portion of the wave power is propagated through the dielectric itself and a part through the surrounding space. The proportionate parts of the electric and magnetic fields resident inside and outside the dielectric are amenable to calculation. As might be expected they depend both on the dielectric constant of the material and on the proximity to cut-off at which the guide is operated. Results of such calculation for the E_0 type of wave are shown in Fig. 3, each for various proximities to cut-off. A dielectric constant of 81 is assumed.

For high dielectric constants and for frequencies far above the cut-off, the power is propagated largely inside the guide whereas for low dielectric constant and for frequencies just above the cut-off, a substantial amount of the power travels outside the guide. In the first case inductive disturbances communicated to neighboring guides are very small and correspondingly the guide is substantially immune to outside disturbances. In the second case these important advantages are absent.

As already stated, many of the properties of wave guides are amenable to calculation. Formulas for the purpose are included in the mathematical paper already referred to. Certain of these properties are intrinsic—as for example, velocity of propagation, attenuation and characteristic impedance. Others may be regarded as extrinsic in that they result largely from the manner in which the guide is used. Examples of the latter are frequency-selectivity and radiation.

Velocity of Propagation

It will be remembered that the velocity of electric waves over ordinary conductors immersed in a particular medium is substantially that of light for that medium. In other words it is equal to the velo-

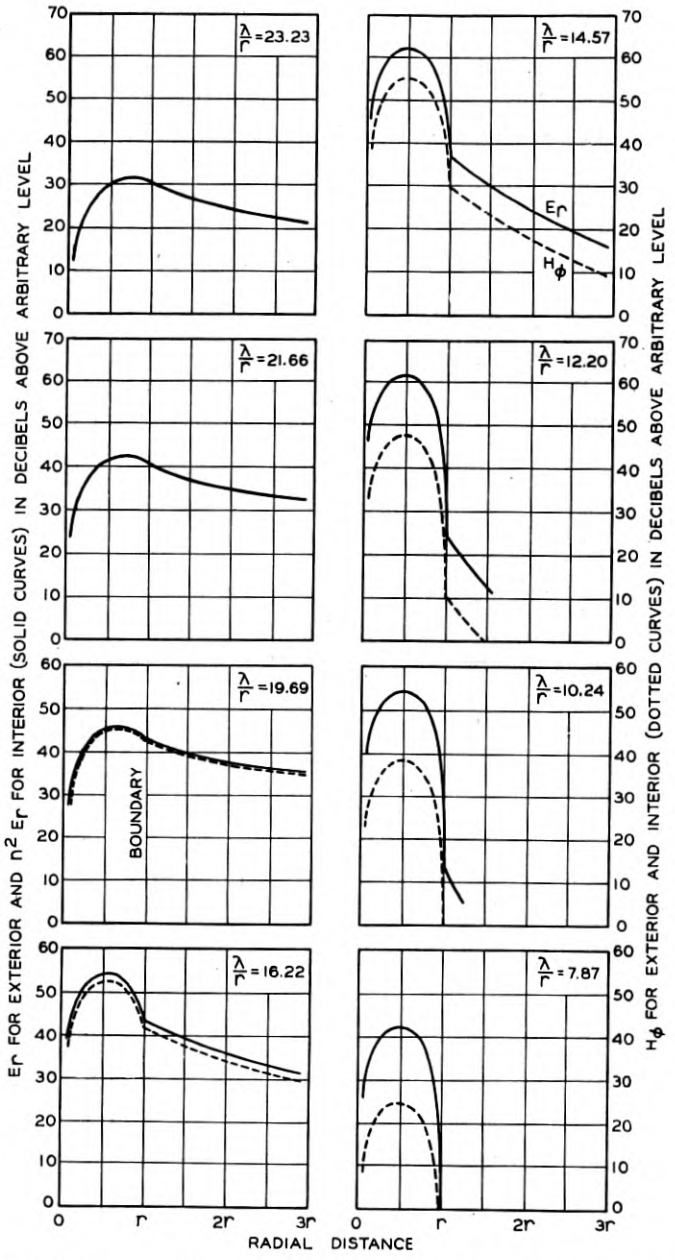


Fig. 3—Relative intensities of electric and magnetic fields inside and outside a dielectric wire while propagating the E_0 type of wave. A dielectric constant of 81 is assumed.

city of light in free space divided by the index of refraction (square root of the dielectric constant). It is also dependent to a small extent on the resistance and permeability of the conductors themselves. The velocity of propagation in wave guides depends not only on these properties but also on the dimensions of the guide as well. For a cylindrical guide it is convenient to express the relation between frequency and dimension as a ratio of wave-length in free space to diameter (λ/d). Also the velocity in the guide may conveniently be expressed as its ratio to the velocity of light ($c/v = k$). Designating by λ the wave-length in free space and by λ_g the wave-length in the guide $k = \lambda/\lambda_g$. Figures 4, 5 and 6 show in graphical form these velocity ratios for three representative cases. The solid curves are calculated. The points are experimental.

Figure 4 covers the case of E_0 waves in a dielectric having a constant

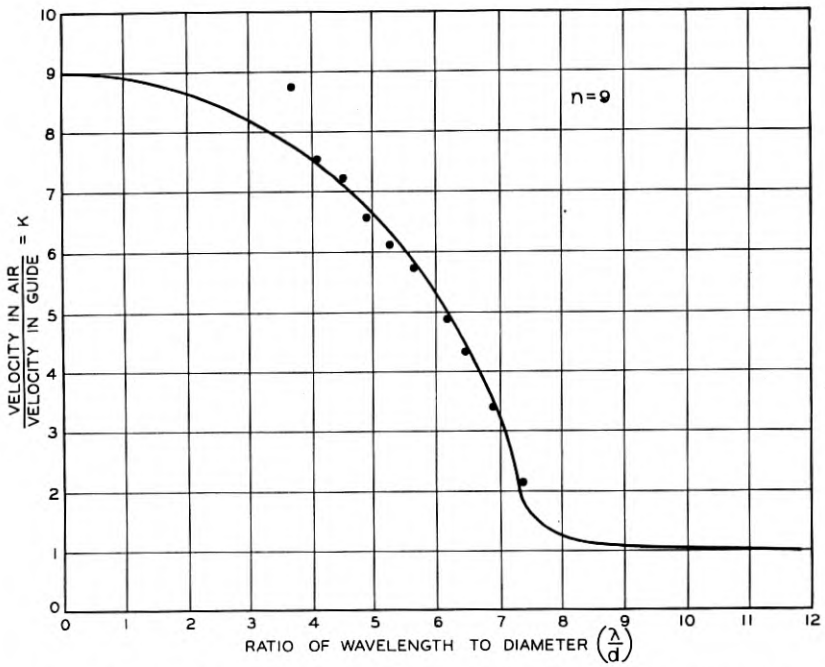


Fig. 4—Velocity ratio for the E_0 type of wave in a dielectric wire ($K = 81$).

of 81 when surrounded by air. It will be observed that at the highest frequencies (lowest values of λ/d) the velocity of propagation is one ninth that of light in free space whereas at the lower frequencies (near cut-off) the velocity is that of light in free space. If the di-

electric constant were progressively lowered the velocity even at the highest frequencies would approach that of light in free space and the curve shown would become progressively flatter. In the limit the dielectric constant would be unity and the velocity ratio also would be unity. Under this circumstance the dielectric wire having a constant substantially the same as that of the surrounding medium would cease to function as a guide.

The experimental points of Fig. 4 were obtained by transmitting waves at each of several frequencies ranging from 100 mc. (λ = 300 cm.) to 400 mc. (λ = 75 cm.) through columns of moderately pure water. The distances between nodes and loops of the standing waves gave data for the velocity of propagation. The method therefore utilized, in a modified form, a technic sometimes invoked for determining the velocity of electric waves on wires or the velocity of sound in air columns. The columns were supported in thin walled bakelite cylinders each about three feet long. Two diameters were used, 6 inches and 10 inches respectively.

Figure 5 covers the case of the same type of waves and the same

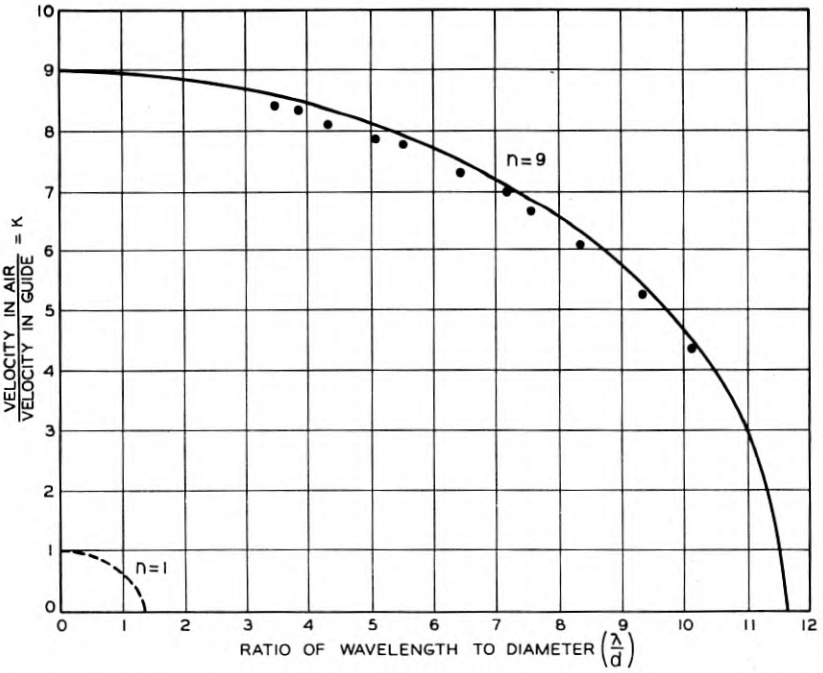


Fig. 5—Velocity ratio for the E_0 type of wave in a metal pipe filled with an insulator (K = 81).

dielectric as above but surrounded by metal. It will be noted that the limits of this curve are essentially the same as for the unshielded guide but that the two curves follow rather different courses. If, in this case, the dielectric constant were progressively reduced the curve shown would gradually shrink into the miniature replica shown dotted in the lower left corner. This is of course the practical case of a hollow conductor to be discussed shortly. The above discussion leads naturally to the view that a wave guide is a propagating medium bounded by a dielectric discontinuity. In one case the discontinuity is the interface between the dielectric of the guide and the surrounding medium. In the other it is the interface between the dielectric and a surrounding conductor.

It will be noted from Fig. 5 that at the highest frequencies the phase velocity in shielded guides, like that for unshielded guides, is the same as the velocity along ordinary conductors in that medium but at frequencies near the cut-off this velocity approaches infinity. The solid curve is calculated on the assumption that the medium had a dielectric constant of 81. The indicated points are the results of experiments made with water as a dielectric. For this experiment the water was supported in three-foot cylinders of copper, six inches and ten inches in diameter respectively. The same range of frequencies was used as above. A somewhat closer agreement between calculation and experiment would have resulted if a value of dielectric constant of 78.9 had been assumed in the computations.

TABLE II

VELOCITY RATIOS FOR H_1 WAVES IN HOLLOW CONDUCTORS

Ratio Space Wavelength to Guide Diameter [1]	Ratio Velocity in Free Space to Velocity in Guide		Difference	Per Cent
	Calculated [3]	Measured [2]		
0.980	0.818	0.818	0.000	0.0
1.033	0.795	0.797	0.002	0.3
1.108	0.757	0.762	0.005	0.7
1.246	0.684	0.683	− 0.001	− 0.1
1.375	0.592	0.601	0.009	1.5
1.469	0.510	0.514	0.004	0.8
1.547	0.424	0.429	0.005	1.2

[1] Probable error 0.4 per cent.
[2] Probable error 0.4 per cent.
[3] Probable errors (arising from error in λ/d) range from 0.2 per cent for $\lambda/d = 0.98$ to 1.8 per cent for $\lambda/d = 1.55$. Note that agreement in most cases is within probable error. However the fact that in all but one observation differences have same sign suggests some systematic relation.

Figure 6 is based on calculations covering the case of a hollow conductor (air dielectric) propagating the H_1 type of wave. The experimental data for plotting points on Fig. 6 were obtained at frequencies extending from 1500 mc. ($\lambda = 20$ cm.) to 2000 mc. ($\lambda = 15$ cm.) on hollow cylinders ranging in diameter from four inches to six inches. Relative velocity was determined from the length of standing waves set up in short sections of these wave guides. The measurements represented were made with much more refined apparatus than utilized in obtaining the data for Figs. 4 and 5.

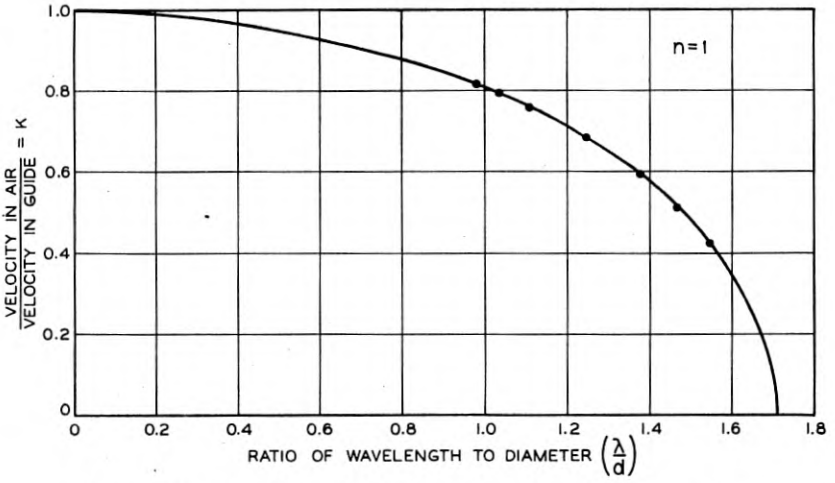


Fig. 6—Velocity ratio for the H_1 type of wave in a hollow metal pipe.

Attenuation

Figure 7 shows in graphical form the calculated attenuations suffered by each of the four more common types of waves when traveling through a hollow copper pipe 5 inches in diameter. It is immediately obvious that the attenuation is infinite for all waves at their respective cut-off frequencies. However, at frequencies above the cut-off this attenuation becomes finite, generally descending to values comparable with attenuations experienced on ordinary conductors at considerably lower frequencies. For the E_0 and E_1 types of waves the attenuation falls from infinity at cut-off to a minimum at a frequency $\sqrt{3}$ times the cut-off frequency after which it again begins to increase and ultimately varies in a linear fashion much as does attenuation over ordinary conductors. For the H_1 type of wave this minimum comes at a frequency 3.15 $\sqrt{3}$ times the cut-off frequency. Thus we see that

for these three types, wave guides are somewhat similar in their behavior to ordinary conductors when operated at the highest frequencies but they depart radically at frequencies near cut-off.

Calculations indicate that the H_0 type of wave has a descending attenuation characteristic at all frequencies above cut-off. This suggests that we may be able to realize very low attenuation merely

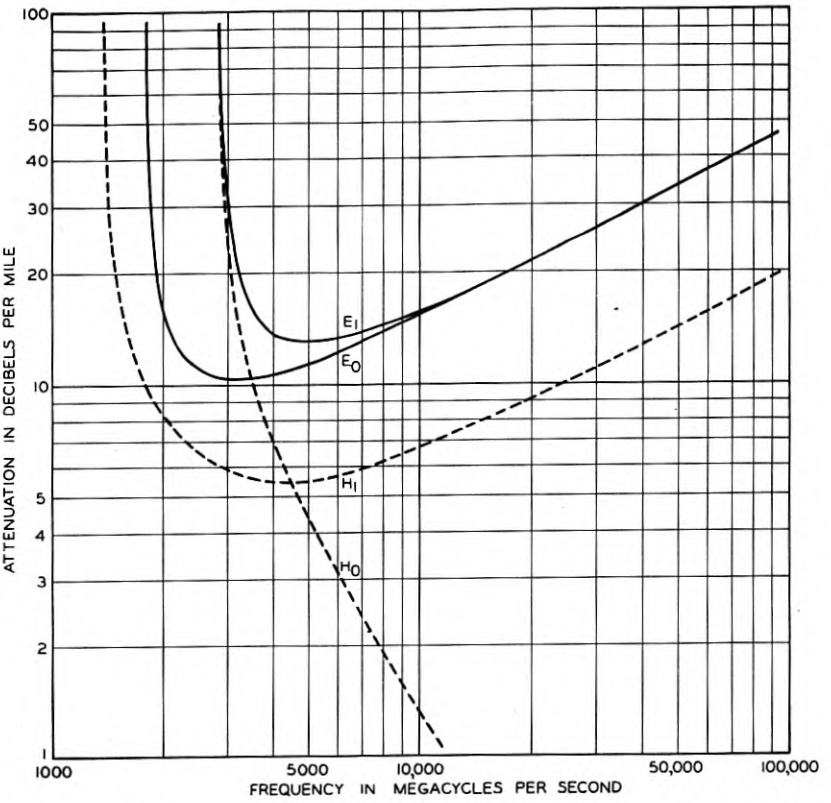


Fig. 7—Attenuations suffered by each of the more common types of waves in a hollow copper pipe 5 inches in diameter.

by raising frequency. This remarkable property is, so far as the author is aware, altogether unique in the realm of electrical transmission. It should be borne in mind, however, that for structures having reasonable dimensions, these low attenuations can only be obtained from frequencies that are above those now readily available. It may be noted in passing that at the minimum of the H_1 curve, transmission is flat to a half db per mile over a band-width of 4000 mc.

The author's experimental work on attenuation is still incomplete, but the results to date are altogether in keeping with calculation. Work done at or near cut-off for all four types of waves confirms their descending characteristics at these points. Other more systematic measurements made on the H_1 type of wave over a considerable range of frequencies are also in good agreement with calculation. Typical results are shown in Fig. 8. They were made on a straight section of

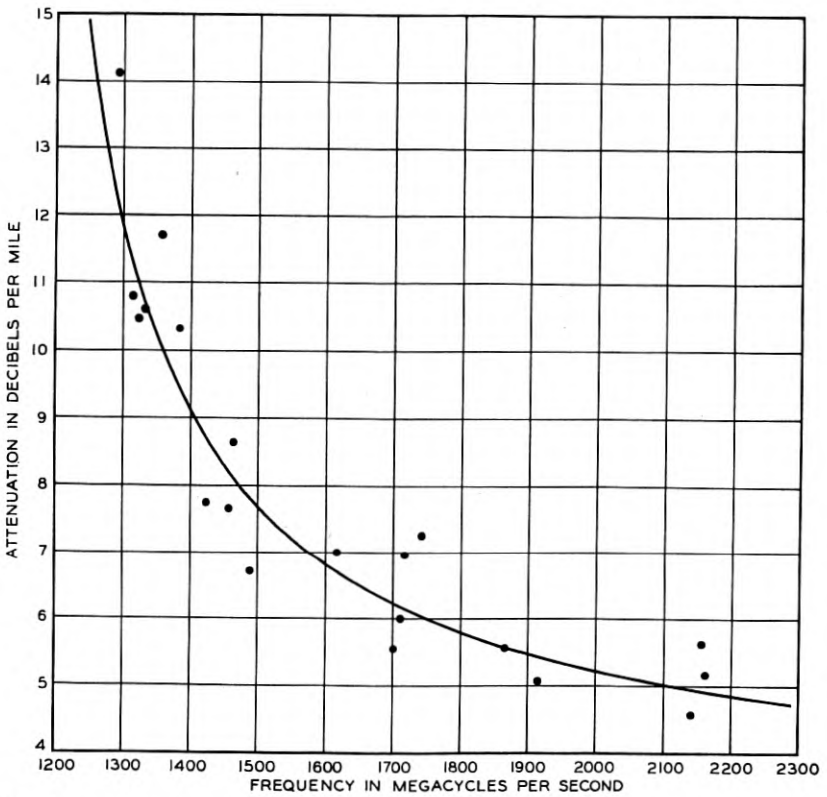


Fig. 8—Attenuation suffered by H_1 waves in a 6-inch hollow copper pipe. Curve is calculated. Plotted points are experimental.

hollow copper pipe six inches in diameter and 1250 feet long. No experimental attenuation data on the H_0 type of wave are yet available except at cut-off. It may be argued, however, that the same theory applies to all four forms of waves so that data tending to confirm the calculated attenuation of one form of wave tends also to substantiate the predicted attenuation for the other forms as well.

Characteristic Impedance

A second intrinsic property of wave guides is characteristic impedance. It may be calculated by integrating the complex Poynting vector over the cross-sectional area and dividing the result by the square of the effective current. Formulas for this purpose are included in the companion mathematical paper referred to above. The numerical results of such calculation are shown in Fig. 9 for a 4-inch diameter hollow copper conductor for each of the four principal waves mentioned above.

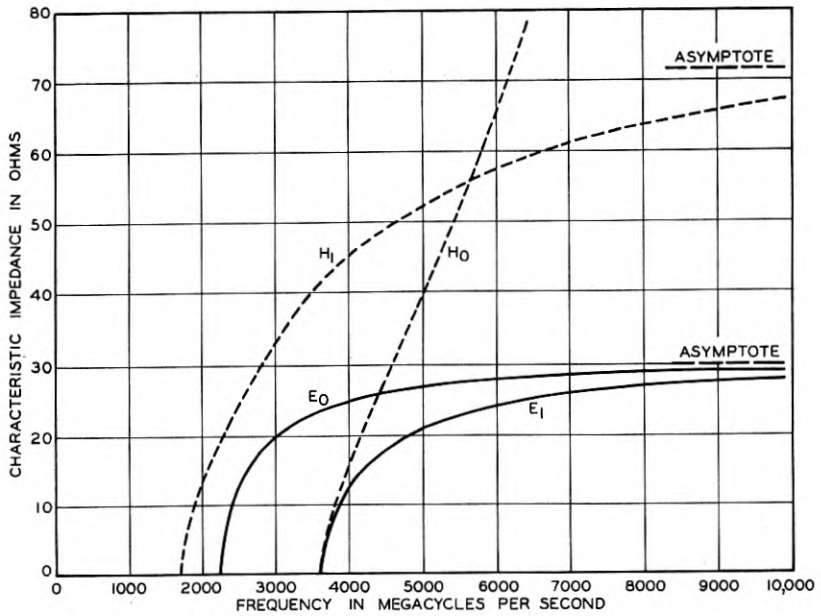


Fig. 9—Calculated values of characteristic impedance of a 4-inch hollow copper pipe for each of the four more common forms of waves.

It will be remembered that when an ordinary wire line is terminated in its own characteristic impedance there are no standing waves. This condition leads to a maximum of power delivered to the receiver. Such an impedance match is sometimes referred to simply as a termination. Terminations for wave guides are entirely similar in their behavior to those of wire lines and may be had by a variety of means. One is a thin film of resistance material placed perpendicular to the axis of the guide followed at a prescribed distance by a perfectly conducting reflector. It is often convenient to provide the latter in

the form of a movable piston. Another form is a resonant chamber containing some dissipative material. Conditions for termination may be calculated in so far as the properties of materials are known or they may be determined experimentally by successive adjustments of film density and piston adjustments until standing waves have been eliminated. Figure 10 shows graphically a typical series of experi-

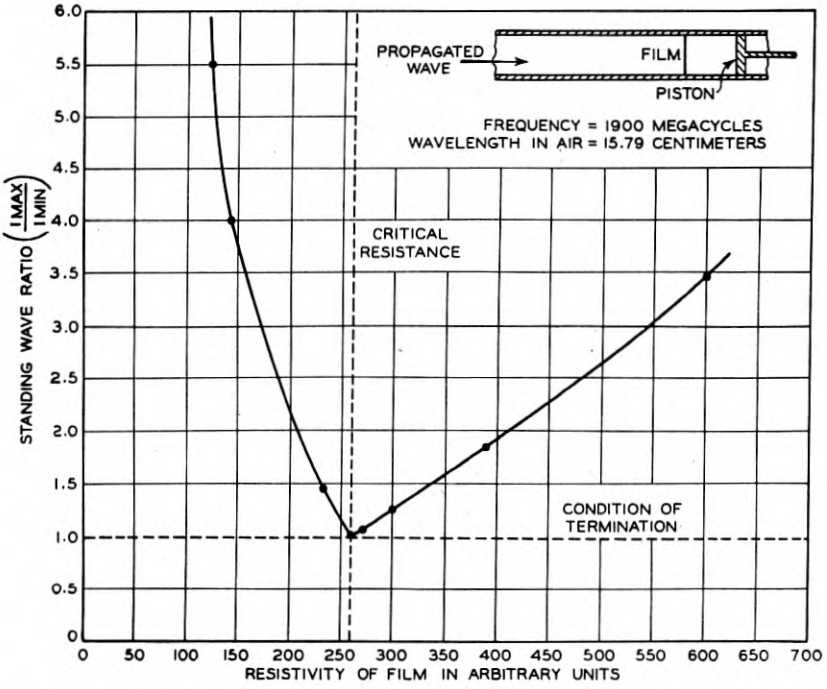


Fig. 10—Typical set of experimental data as various degrees of impedance match are obtained.

mental data of the magnitudes of standing waves as various degrees of impedance match are obtained.

Frequency Selectivity

It is evident from Fig. 7 above that wave guides are inherently high-pass filters. There is still another property of a wave guide that may also provide selectivity. It depends on the principle of standing waves. By this means, resonance effects may be produced that make a short section of guide behave somewhat as if it were a simple series circuit consisting of an inductance and a capacity in series

with an electromotive force. Under other conditions, it may behave as a circuit made up of inductance and capacity in parallel with an electromotive force.[2] At still other frequencies it may present to a source a positive (inductive) reactance or a negative (capacitive) reactance. This makes possible circuit elements which may be combined to form various filter or network equivalents. We may have, therefore, from wave guides frequency selection by either or both of two fundamentally different properties.

Radiation

Discontinuities in wave guides, particularly those in which no shield is present, tend toward losses by radiation. In the case of a hollow conducting pipe radiation issues from the open end much the same as sound waves from a hollow tube. It has been possible to expand the ends of these pipes into horns, thereby obtaining effects very similar to those common in acoustics. Such an electrical horn not only possesses considerable directivity but it may also provide a moderately good termination for the pipe to which it is connected. In so doing its function is probably quite analogous to that of a true acoustic horn which provides an efficient radiating load for its sound motor.

Some Apparatus and Methods Used in Wave Guide Studies

It is obvious, of course, from the very nature of guided waves that the apparatus and methods must be rather different from the more common electrical methods. This difference is such that an adequate description would require more space than is here available. However, for purposes of completeness a few of the more interesting and fundamental aspects of the experimental side are included below. For the most part this description will center around the H_1 type of wave. (See Fig. 1 above.)

The Simple Resonant Chamber

In much the same way that the simple tuned circuit containing localized inductance and capacity is fundamental to the radio art so also is the simple resonant cavity fundamental to wave guide work. Although it may assume a variety of forms, one of the more obvious is a short piece of cylindrical wave guide, preferably of hollow metal pipe bounded by a piston and an iris diaphragm as shown in Fig. 11.

[2] In pursuing this work it has been convenient at times to refer not only to *circuit* analogues but also to *optical* and also *acoustical* analogues. This has been due in part to the lack as yet of an adequate vocabulary and in part to the hybrid nature of the subject at hand.

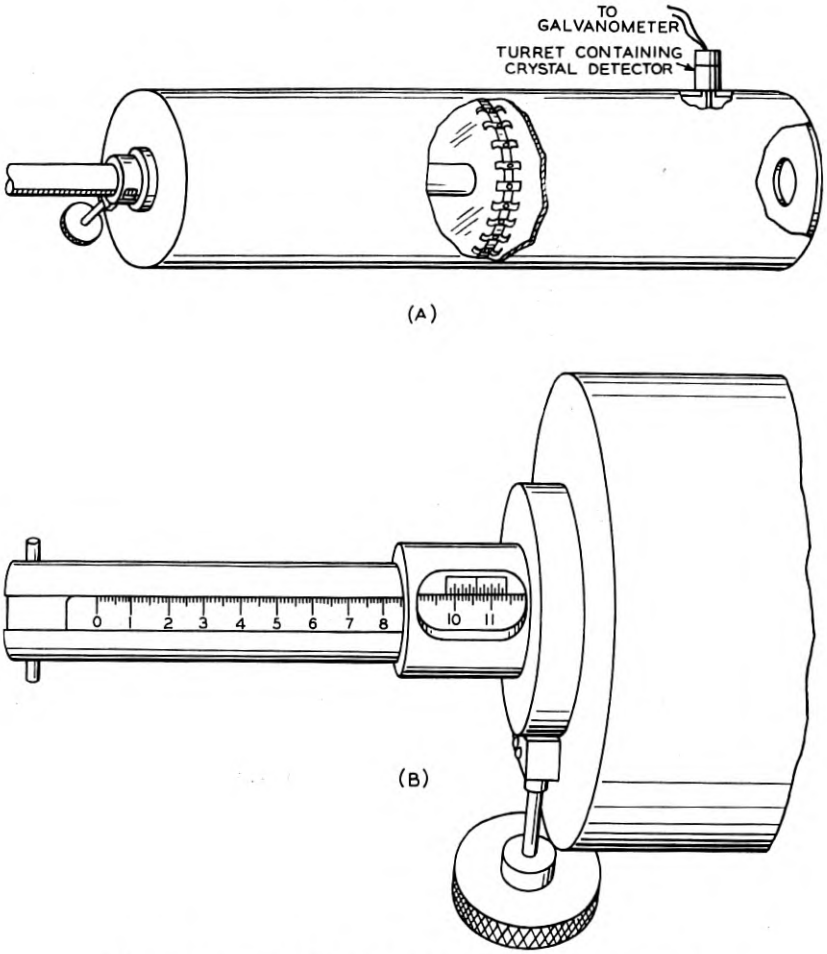


Fig. 11—A tunable chamber resonant to short electric waves.

In its role as a tuned circuit, the resonant chamber is sometimes used as a wave-meter, sometimes in connection with a generator of short waves (thereby enabling a vacum tube to work more effectively) and sometimes as an element in a receiver (thereby impressing on a detector a maximum of the received wave power). When such a chamber is excited by very short electric waves and is varied in length, resonance takes place at certain specified intervals depending on the frequency and phase velocity. This condition may be detected either by a crystal detector and meter located just outside of the iris opening or by a somewhat more elaborate arrangement whereby a crystal

mounted in a shielded cartridge is coupled to the chamber by a probe wire perhaps a quarter inch long extending through a hole in the wall. Fig. 11*A* shows one form of resonant chamber complete with detector. The piston position is read off on a scale and vernier (Fig. 11*B*). Successive positions at which resonance is noted give data for determining velocity of propagation. Chambers of this kind having various diameters were used to verify the velocity ratios shown earlier in this paper. Electrical connection between piston and walls may be had by numerous phosphor bronze fingers, or perhaps by ball bearings distributed in a race around the periphery. Good contact is not always essential. In fact, fair work may sometimes be done with a loosely fitting piston or even an insulated piston.

Resonant chambers may be activated merely by placing them within a foot or two of a source of waves such as a Barkhausen oscillator. Their dimensions must, of course, conform to the wave-length requirements as outlined above. Standard 5-inch OD brass pipe having one-sixteenth inch wall has been found satisfactory for the frequency range from 1500 mc. ($\lambda = 20$ cm.) to 2000 mc. ($\lambda = 15$ cm.). Any convenient length around 2 feet is appropriate for the variable type of chamber.

Generators

One arrangement for generating the H_1 type of wave consists of connecting the primary source of waves between diametrically opposite points on the inside of a hollow cylindrical conductor as shown by Fig. 12*A*. This primary source may consist of a positive grid (Barkhausen) tube or a magnetron.[3] Both have been used successfully to give frequencies up to about 3330 mc. ($\lambda = 9$ cm.).

A typical arrangement of such an oscillator is shown in Fig. 12*B*. The terminals of the spiral grid of the Barkhausen tube are connected to diametrically opposite points through a suitable by-pass condenser. The filament and plate leads enter along a plane perpendicular to that of the grid. Since the grid leads correspond to lines of electric force in the generated wave, the diametral plane perpendicular thereto corresponds to an equipotential. By locating the plate and filament leads in such an equipotential, their presence will not materially affect the normal field prevailing in the chamber. In the design shown the filament connectors constitute the outside plates of a three-plate by-pass condenser. The third or central plate is a rigid member grounded on the main guide. It connects to the plate of the Barkhausen tube. Connections to the exterior are had through five

[3] "Vacuum Tubes as High-Frequency Oscillators," M. J. Kelly and A. L. Samuel, *B.S.T.J.*, Vol. 14, p. 97, January 1935.

insulated binding posts. The oscillator unit shown carries on its exterior a plug connector leading by cable to a nearby d-c. power supply unit.

If an oscillator similar to that described above were connected into the middle of a long hollow pipe, waves, would of course, be propagated in both directions. Those that would ordinarily be propagated to the left may be reflected by a suitably located reflecting wall or piston so

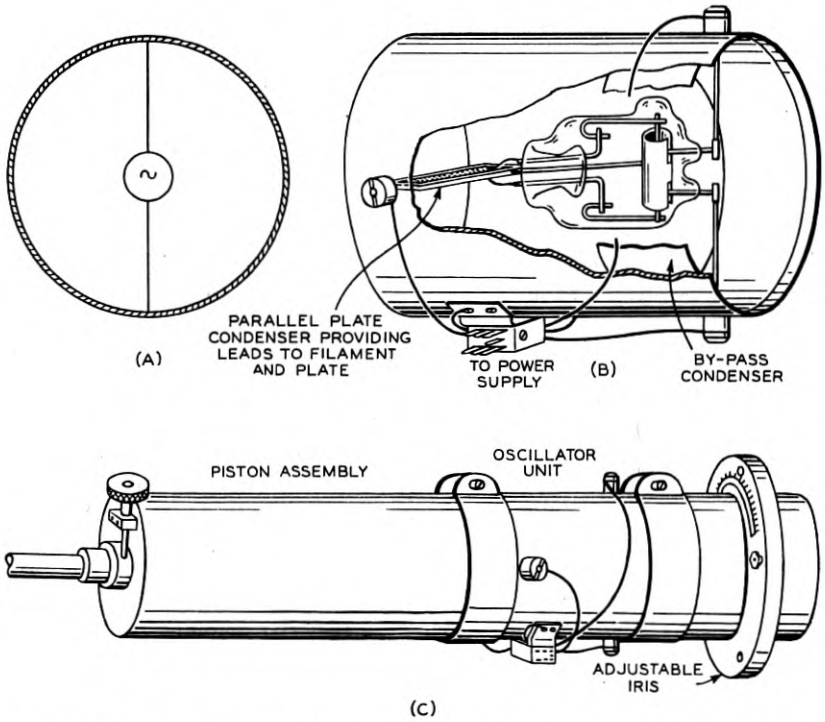


Fig. 12—Various component parts of a wave guide generator. (*A*) Schematic representation. (*B*) The oscillator unit. (*C*) Complete generator including oscillator piston and iris.

as to reinforce those being propagated to the right. Also an iris of suitable proportions may be so located in front of the generator as to further enhance oscillations. As has been pointed out above the section of pipe bounded by the piston and iris together approximate in behavior a tuned circuit. It is convenient to regard the chamber as a load impedance characteristic of the tube itself or perhaps it should more properly be regarded as a transformer by which the oscillator is matched to the line.

In practice the generator may conveniently be built up from an oscillator unit, a piston assembly and an adjustable iris, all of the same diameter of pipe fastened together by exterior metal clamps as shown in Fig. 12*C*. The open end of this generator may be connected to a guide over which transmission is desired or it may be coupled loosely to some nearby laboratory apparatus on which measurements are to be made.

The total length of the chamber and hence the piston setting will of course depend on the frequency to be generated. In general this will be roughly an integral number of half wave-lengths. The relative position of the oscillator along the length of the chamber will depend on its impedance characteristics and to some extent on the diameter of the iris opening. For a piece of laboratory apparatus where frequency variability is desired these various dimensions should preferably be adjustable as shown. If a source of single frequency is desired, the resulting apparatus may be greatly simplified as all of these dimensions may be fixed at the time of construction.

The Tuned Receiver

By reversing the principle used in the generator above, replacing the oscillatory source by a suitable indicator the resonant chamber becomes effectively a simple tuned receiver. If the indicator is appropriately located along the length of the chamber, substantially all of the incident power will be absorbed and the device as a whole will be a veritable sink of wave power. It may be clamped to the end of a long wave guide, thereby constituting a termination, or it may be used to pick up short radio waves of not too small amplitude. See Fig. 13.

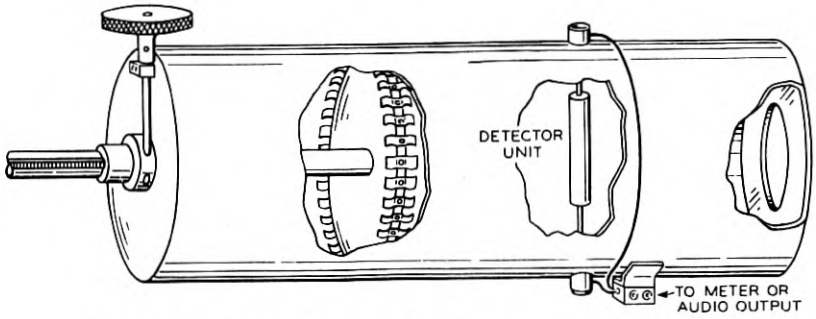


Fig. 13—A tuned receiver based on the resonant cavity principle.

Indicators

It is often desirable to have available in the laboratory some kind of a wave indicator or probe such as shown in Fig. 14. This one consists of a simple silicon detector in cartridge form, together with a microammeter, both mounted on a fibre support of convenient size and shape for exploring the fields prevailing around any piece of apparatus. It is easy to show by this means that there are no appreciable fields prevailing around a generator such as described above except near the orifice. Also this probe may be used to determine the

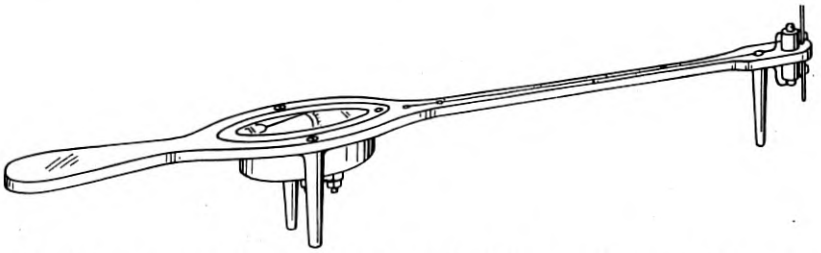


Fig. 14—A convenient probe for exploring the field around a source of waves.

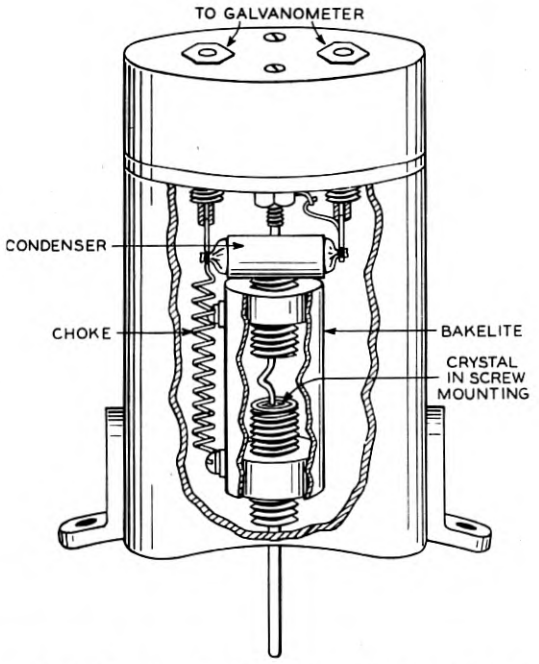


Fig. 15—A detector mounting suitable for indicating the presence of waves in a guide.

approximate orientation of the lines of electric force in the wave front as well as the general directive pattern of the radiation.

Figure 15 shows details of a crystal detector mounting suitable for an indicator on a wave-meter. When silicon crystals are used, units may be had that will hold their calibrations moderately well over considerable periods of time. Thermocouples of both the cross wire and deposited type have been used with moderate success. Also diode and triode rectifiers have been tried. However, for general laboratory use where simplicity and convenience are important the crystal detector is perhaps best.

Wave-Meters

It is, of course, desirable in this work to know the frequency or wave-length being used. The simple resonant chamber already described enables wave-length to be measured accurately. However, such a device does not give directly the wave-length in free space since in these chambers phase velocities are in general greater than ordinary light. It is true, of course, that a suitable conversion curve can be prepared. However, it is often more convenient to use for a wave-meter some form of a coaxial conductor system on which standing waves may be measured. These will be very nearly at least the length of the corresponding waves in free space.

Figure 16 shows a wave-meter based on this principle. The conductor (*a*) and the hollow cylinder(*b*) constitute the coaxial conductors. A bridge (*c*) in the form of a conducting disc is made movable by means of the threaded tube (*d*) which passes over the central conductor (*a*). This tube carries a millimeter thread engaged by the knurled head (*e*). One complete turn of this head therefore raises or lowers the bridge by one millimeter. If coarse adjustments are desired the head may be disengaged from the threaded tube by a cam operated by the knob (*f*), and the tube raised or lowered by taking hold of its extended portion. The outer conductor or shell carries an open slot (*g*) through which an index (*h*) attached to the shorting bridge (*c*) extends. This index passes over a centimeter scale (*i*). The outer conductor is mounted on a short piece of wave guide so that the apparatus may be clamped in line with other apparatus. The inner conductor (*a*) extends through a small opening in the section of wave guide far enough to extract from the passing waves enough power for activating the wave-meter. This coupling may be varied as needed by extending or retracting a third small rod (*jj'*) running through the center of central conductor (*a*).

Resonance is indicated by a crystal detector (*k*) and d-c. meter. This detector is only loosely coupled to the coaxial system by a small

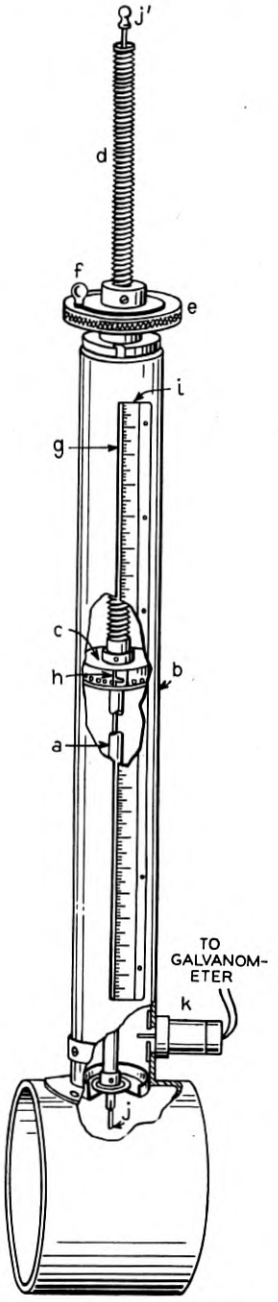


Fig. 16—A form of coaxial conductor wave-meter particularly adaptable to wave guide work.

pick-up wire extending through the walls of the hollow cylinder. This form of wave-meter is moderately fast and permits wave-length differences of one or two hundredths of a centimeter to be readily detected.

Miscellaneous Apparatus

Sometimes it is desirable to change the length of a pipe without changing its diameter. For this purpose telescoping pipe is to be avoided. A pipe with removable sections may, however, be provided. Units of 10 cm., 5 cm., 2 cm., 2 cm. and 1 cm. have been found convenient. They are aligned in a slightly larger half-section of the same kind of pipe which provides their support.

It may be desirable at times to investigate the field inside a pipe to determine if standing waves are present. This may be done by mounting a detector similar to that shown in Fig. 15, on a carriage so it may be advanced along a slot cut in a piece of wave guide perhaps 60 cm. long. Often it is necessary to pass from one size of pipe to another. A conical reducer perhaps 30 cm. long may be used for this purpose.

It is usually desirable to construct components such as the above with lengths of some integral number of centimeters such as 10 cm., 20 cm. or 50 cm. This obviously facilitates the addition of the component lengths used and often simplifies calculation.

It is obvious from the above that a laboratory working with wave guides must use for its circuit components such unusual electrical items as hollow pipes, movable pistons and iris diaphragms. These should be capable of quick assembly into a variety of forms, sometimes as a generator, sometimes as a tuned receiver and sometimes as a termination. This object imposes a wide range of requirements that can best be met by mounting the parts by means of clamp supports on a saw-horse arrangement or wave guide bench such as shown in Fig. 17.

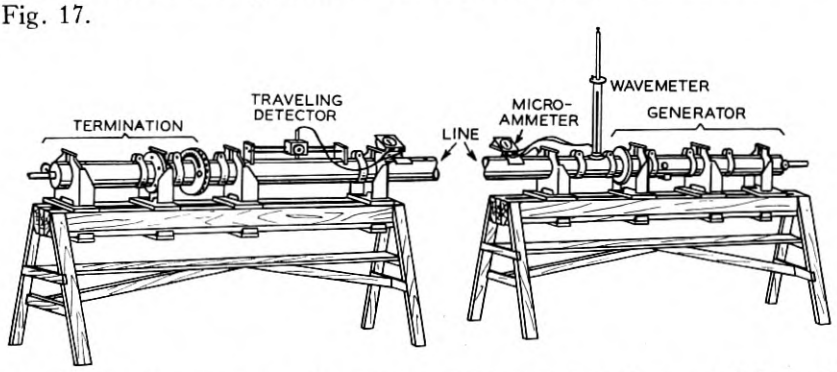


Fig. 17—Bench mountings with typical apparatus used at the transmitting and receiving ends of an experimental wave guide.

References

1. "On the Passage of Electric Wave Through Tubes, or the Vibrations of Dielectric Cylinders," Lord Rayleigh, *Phil. Mag.*, Vol. 43, 1897, pp. 125–132.
2. "Electromagnetic Waves in Dielectric Wires," Hondros and Debye, *Ann. d. Phys.*, Vol. 32, 1910.
3. "Detection of Electromagnetic Waves on Dielectric Wires," Zahn, *Ann. d. Phys.*, Vol. 49, 1916.
4. "Electromagnetic Waves in Dielectric Conductors," Schriever, *Ann. d. Phys.*, Vol. 63, 1920.
5. French Patent 757,972, October 23, 1933.
6. British Patent 420,447, December 3, 1934.
7. "Measurements in the Radiation Field of a Linear Antenna Excited Inside a Hollow Cylinder," Bergmann and Krügel, *Annalen der Physik*, Vol. 21, No. 2, page 113, October 29, 1934.
8. A Paper entitled "Transmission of Electromagnetic Waves Through Hollow Tubes of Metal," by W. L. Barrow of Massachusetts Institute of Technology, is in an advanced stage of preparation. It is intended for presentation before the Washington Meeting of the U.R.S.I. and I.R.E., May 1, 1936, and for publication in an early issue of the *Proceedings of the Institute of Radio Engineers.*

Hyper-Frequency Wave Guides—Mathematical Theory

By JOHN R. CARSON, SALLIE P. MEAD and S. A. SCHELKUNOFF

Following a brief historical sketch, this paper deals with the mathematical theory of wave transmission in two novel kinds of cylindrical wave guides of circular cross section; namely, the hollow conductor and the dielectric wire. These transmission systems behave as high pass filters with exceedingly high critical frequencies.

The attenuation and impedance characteristics of the hollow conductor, heretofore ignored as far as the writers are aware, are given especial attention. This investigation discloses the remarkable fact that there exists in this system *one* and *only one* type of wave, the attenuation of which *decreases* with increasing frequency, a characteristic which attaches to no other type of guided wave known to the writers.

I. Introduction

THE object of this paper is to derive and discuss the characteristics of two novel guided wave *transmission* systems. Structurally one consists simply of a straight hollow [1] conducting cylinder of circular cross-section. The electromagnetic wave is confined inside the cylindrical sheath and is propagated along the axis of the cylinder. The other consists simply of a dielectric wire, within which the major part of the electric field is confined. The mathematical theory developed below does not deal with the question as to how such waves are established nor with the reflection phenomena which must occur at the terminals and other points of discontinuity. The analysis is limited to finding the types of waves which are possible in such systems, and to investigating and describing their characteristics.

The historical background of the problem is interesting. In 1897 Rayleigh published a paper entitled "On the Passage of Electric Waves through Tubes, or the Vibrations of Dielectric Cylinders." [2] Dealing solely with ideal cylinders of perfect conductivity he showed that for all types of waves that can exist inside the cylinders there are critical frequencies below which the waves are attenuated and above which they are freely transmitted. The first paper on transmission along dielectric wires was that published in 1910 by Hondros and Debye entitled "Elektromagnetische Wellen an dielektrischen Drähten." [3] This deals theoretically with transmission along cylinders of ideally non-conducting material, somewhat along the lines followed in Section IV

[1] The term *hollow* means that the interior of the cylinder is electrically non-conducting.

[2] *Phil. Mag.*, Vol. 43, 1897, pp. 125–132.

[3] *Ann. der Phys.*, Vol 32, 1910, pp. 465–476.

of this paper. Another paper, entitled "Über den Nachweis elektromagnetischer Wellen an dielektrischen Drähten," [4] published in 1916 by Zahn, is of interest because of the historical note attached, which indicates that experimental work was begun in 1914 by Rüter and Schriever, two students of Zahn, and continued with such diligence as the exigencies of war permitted until the date of Zahn's paper, at least. In 1920 Southworth, then working at Yale University, accidentally observed such waves in a trough of water which he was using in connection with some high-frequency studies, measured their wave-lengths and recognized their identity with those discussed by Schriever [5] in a paper which appeared at about that time. In 1924 Carson rediscovered the transmission characteristics of the hollow conducting cylinder, and disclosed it in an unpublished memorandum entitled "Hyper-Frequency Wave Filters." Finally, in 1931, Southworth, then a research engineer with the American Telephone and Telegraph Company, returned to the subject and initiated the comprehensive investigation which he is reporting in a companion paper.[6] Independently, and almost simultaneously, Hartley, at the Bell Telephone Laboratories, suggested the possibility of guided transmission along a *hollow* cylindrical dielectric wire; and these two (Southworth and Hartley) enlisted our cooperation in a mathematical investigation.

In the theoretical parts of these papers dissipation was always neglected, though obviously the attenuation would be a controlling factor in practical applications. The writers, on the other hand, have given especial attention to this factor. Out of this research there emerged the remarkable fact that with hollow conducting guides there exists *one* and *only one* type of wave the attenuation of which *decreases* with increasing frequency; a unique characteristic which does not attach to dielectric wires, nor so far as the writers are aware, to any other type of guided wave.

IA. Transmission Through Hollow Conducting Cylinders

Throughout this paper it will be assumed that the cylindrical sheath possesses high conductivity and that the losses in the internal dielectric medium are either small or negligible. Subject to these assumptions the effect of dissipation on the attenuation of the wave is formulated in

[4] *Ann. der Phys.*, Vol. 49, 1916, pp. 907–933. This paper contains several collateral references.

[5] "Elektromagnetischen Wellen an Dielektrischen Drähten," *Ann. der Phys.*, Vol. 63, 1920, pp. 645–673.

[6] "Hyper-Frequency Wave Guides—General Considerations and Experimental Results," G. C. Southworth, this issue of the *Bell System Technical Journal*.

Section III. First, however, in the general discussion which immediately follows and, in particular, in the comparison with the usual guided wave transmission systems, attention will be confined to the ideal non-dissipative structure. This simplification brings out, in a simpler and more striking way, the peculiar transmission characteristics of the system, while, at the very high frequencies involved, it introduces negligible error except as regards the attenuation due to dissipation.

In the ordinary type of guided-wave systems, such for example, as that composed of two concentric conductors, or two parallel wires, the guiding conductors form two sides of a *circuit* in which equal and opposite currents flow, and the transverse lines of electric intensity terminate on the two sides of the circuit. In the system under consideration there is only *one* conductor and consequently there is no *circuit* in the usual sense. Corresponding to this difference in physical structure there are striking differences in the character of the waves propagated.

In the first place, in the ordinary type of guided wave system, the wave employed for the transmission of power and intelligence is the *Principal Plane Wave.* For the ideal non-dissipative case, the field of this wave is entirely transverse to the axis of the system; that is, the axial components of the electric and magnetic intensities are everywhere zero. Furthermore all frequencies are transmitted without attenuation with the same phase velocity; that of light in the medium. (Of course dissipation modifies the phenomena somewhat but in actual systems designed for efficient transmission the Principal Wave approximates to that just described.)

In the hollow conducting cylinder, on the other hand, *no principal transverse wave can exist;* that is, there must exist inside the cylinder either an axial component of the electric or the magnetic intensity, or both. Physically this is answerable to the absence of a circuit on which the transverse lines of force might terminate. Thus in the hollow conducting cylinder all the possible waves must be *complementary* waves;[7] a type which is ignored in the ordinary transmission system.

A second outstanding distinction is that in the hollow conducting cylinder, all frequencies below a critical frequency are attenuated while frequencies above the critical frequency are freely transmitted without attenuation.[8] In this respect the system behaves like a Campbell high-

[7] See "Guided and Radiated Energy in Wire Transmission," John R. Carson, *Jour. A.I.E.E.*, October 1924.

[8] It will be understood, of course, that this is strictly true only in the ideal case of no dissipation.

pass wave-filter. The exact value of the critical frequency depends, as shown later, on the type of wave transmitted; roughly speaking, however, the internal diameter must be approximately equal to one-half a wave-length in the internal dielectric medium at the lowest critical frequency. (The exact formula is diameter $> \frac{3.68}{2\pi}$ times the wave-length.) Since we are interested in freely transmitted waves it is evident at once that for a cylinder of practicable dimensions the frequencies employed must be relatively enormous. For this reason it may be appropriately said that the hollow conducting cylinder is applicable to the transmission of *hyper-frequency* waves alone.

The types of waves which can exist inside the cylinder are broadly classifiable as E-waves and H-waves.[9] By the term E-wave is to be understood a wave in which the axial component of the magnetic force is everywhere absent; correspondingly in the H-wave the axial component of the electric force is everywhere absent. In the E-waves the surface currents in the cylinder are entirely parallel to the axis thereof. On the other hand, in the H-waves the currents may have both transverse and axial components; that is, circulatory components around the periphery of the cylinder in planes normal to its axis as well as components parallel thereto.

In each class of wave there may exist a fundamental wave and in addition *geometrically* harmonic[10] waves. In the fundamental wave the phenomena do not vary around the periphery of the cylinder. In the nth harmonic wave (E_n- or H_n-wave) the phenomena vary around the periphery as $\cos n(\theta - \theta_n)$.

Each component E- or H-wave has its own individual critical frequency. Curiously enough the lowest critical frequency is possessed by the first harmonic H-wave; that is the H_1-wave. For this wave the critical frequency is given by the formula $d > \frac{3.68}{2\pi}\lambda$ where d is the internal diameter of the cylinder and λ the wave-length. In general, however, the critical frequency increases with the order of the harmonic.

In the usual transmission system, the transmission phenomena are determined and described in terms of the characteristic impedance and the propagation constant. By characteristic impedance the engineer understands the impedance actually presented by an infinitely long line to an electromotive force connected across the terminals of the circuit. Now since in the hollow conducting cylinder there is only one

[9] This terminology has been adopted as a matter of convenience. It is suggested by equations (1) where the field is expressed in terms of E_z and H_z. Another terminology is *transverse magnetic* and *transverse electric* waves.

[10] This term must not be confused with *frequency* harmonics.

conductor and hence no circuit, this concept breaks down. There is another way, however, in which the characteristic impedance may be defined, and by aid of which it remains a useful concept in hollow cylinder transmission. Writing $K = K_R + iK_I$ as the complex expression for the characteristic impedance, then it may be shown that

$$K = \overline{W} + i2\omega(\overline{T} - \overline{U}),$$

where $\overline{W}$ is the mean power transmitted, $\overline{T}$ is the mean stored magnetic energy, and $\overline{U}$ the mean stored electric energy, corresponding to an unit current. Now in the hollow conducting cylinder, for, say the E_0-wave, we can calculate

$$\overline{W} + i2\omega(\overline{T} - \overline{U})$$

for an *unit axial* current, and call this the characteristic impedance. Again for the H_0-wave we can calculate this quantity for an *unit circulating current per unit length* and designate it as the characteristic impedance. In addition, somewhat similar conventions apply to the harmonic waves.

One of the chief uses of the foregoing concept of characteristic impedance is in the calculation of the attenuation in the dissipative system. For, if corresponding to $\overline{W}$ we calculate the mean dissipation $\overline{Q}$ per unit length, then the attenuation α is given by

$$\alpha = \overline{Q}/2\overline{W}.$$

All actual systems are of course dissipative and consequently the wave is attenuated. If the hollow conducting cylinder were to be employed in practice for hyper-frequency wave transmission the securing of low and desirable attenuation characteristics would probably be the controlling consideration.

The attenuation in the free transmission range is due to (1) dissipation in the cylinder or sheath and (2) dissipation in the internal dielectric medium. The former is inherent and can be reduced only by employing a sheath of high conductivity and by properly designing the dimensions of the system. As regards the dielectric loss, this may be substantially eliminated by employing air as the dielectric medium. The use of a dielectric medium of high specific inductive capacity has the advantage of substantially reducing the critical frequency; on the other hand it inevitably introduces heavy losses and thus sharply increases the attenuation. The analysis of Section III brings out the remarkable fact that for the fundamental H-wave the attenuation decreases with increasing frequency; for all the other types it increases.

For the very high frequencies with which we shall be concerned in the following analysis, a physically very thin cylindrical metallic sheath behaves electrically as though it were infinitely thick. This fact greatly simplifies the mathematical treatment; its real importance, however, is that external interference is entirely eliminated.

As stated at the outset, this paper will not attempt to deal with the problem of the reflection phenomena which occur at the terminals of the system and at points of discontinuity. For a discussion of the general character of the boundary problem the reader is referred to "Guided and Radiated Energy in Wire Transmission."[7] It may be remarked here, however, that the simple engineering boundary conditions (continuity of current and potential) are entirely inadequate.

IB. Transmission Through Dielectric Guides

The greater part of this paper deals with transmission in thin hollow conducting cylinders; the last section, however, discusses briefly transmission along the dielectric wire.[3] Theoretically this type of transmission is extremely interesting and the mathematical theory resembles to a considerable extent that of hollow cylinder transmission. Unfortunately, however, dielectric losses are usually high. Hence our discussion of dielectric waves will be limited to a development of the fundamental equation from which the critical frequencies and the phase velocities can be determined.

II. Non-Dissipative Hollow Conducting Guides

In dealing with the propagation of hyper-frequency electromagnetic waves inside a long hollow conducting cylinder parallel to the z-axis, it is convenient to write the field equations in the appropriate cylindrical coordinates (ρ, θ, z) in the form,[11]

$$\begin{aligned}
\lambda^2 H_\rho &= \frac{h^2}{\mu i\omega}\frac{1}{\rho}\frac{\partial}{\partial\theta}E_z - \gamma\frac{\partial}{\partial\rho}H_z,\\
\lambda^2 H_\theta &= -\frac{h^2}{\mu i\omega}\frac{\partial}{\partial\rho}E_z - \frac{\gamma}{\rho}\frac{\partial}{\partial\theta}H_z,\\
\lambda^2 E_\rho &= -\gamma\frac{\partial}{\partial\rho}E_z - \frac{\mu i\omega}{\rho}\frac{\partial}{\partial\theta}H_z, \qquad (1)\\
\lambda^2 E_\theta &= -\frac{\gamma}{\rho}\frac{\partial}{\partial\theta}E_z + \mu i\omega\frac{\partial}{\partial\rho}H_z,\\
\operatorname{div} E &= 0, \quad \operatorname{div} H = 0.
\end{aligned}$$

[11] In this form the field is expressed explicitly in terms of the axial electric and magnetic intensities and their spatial derivatives. This is highly advantageous for the purposes of this paper.

In these equations the symbols have the following significance:

E_ρ, E_θ, E_z = components of electric force,
H_ρ, H_θ, H_z = components of magnetic force,
$\lambda^2 = \gamma^2 - h^2$,
γ = propagation constant,
$h^2 = \mu i\omega(4\pi\sigma + \epsilon i\omega/c^2) = 4\pi\sigma\mu i\omega - (\omega^2/v^2)$,
$v = c/\sqrt{\epsilon\mu}$ = velocity of light in the medium,
c = velocity of light in air,
μ = permeability of the medium in electromagnetic units,
σ = conductivity of the medium in electromagnetic units,
ϵ = dielectric constant of the medium in electrostatic units,
$\omega/2\pi$ = frequency,
$i = \sqrt{-1}$.

The solutions of these equations for the axial components of electric and magnetic force, E_z and H_z respectively, in the region, $0 \leqq \rho \leqq a$, a being the internal radius of the conductor, are of the form

$$\begin{aligned} E_z &= \sum_{n=0}^{\infty} J_n(\rho\lambda)(A_n \cos n\theta + B_n \sin n\theta) \exp.\ (i\omega t \pm \gamma z), \\ H_z &= \sum_{n=0}^{\infty} J_n(\rho\lambda)(C_n \cos n\theta + D_n \sin n\theta) \exp.\ (i\omega t \pm \gamma z), \end{aligned} \tag{2}$$

where A_n, B_n, C_n and D_n are arbitrary constants to be determined by boundary conditions and J_n is the Bessel function of the first kind or the internal Bessel function. The components of the transverse electromagnetic field may then be expressed by introducing (2) in (1).

We shall first discuss the simplest case, that in which there is no dissipation. The current will then be in a sheet on the surface, $\rho = a$, of the perfectly conducting cylinder. But the axial current density u_z and the circulating current density u_θ are given by

$$u_z = \frac{1}{4\pi} H_\theta, \quad \rho = a \tag{3}$$

and

$$u_\theta = \frac{1}{4\pi} H_z, \quad \rho = a. \tag{4}$$

Thus it follows that H_z and H_θ are discontinuous at the surface $\rho = a$ and the boundary conditions are simply $E_z = E_\theta = 0$. These conditions can be fulfilled by two types of waves: (1) a wave for which H_z

is zero everywhere, which will be called generically the E-wave and (2) a wave for which E_z is zero everywhere, which will be designated generically as the H-wave. (If the cylinder is dissipative, however, the E- and H-waves can exist alone only for the case of circular symmetry. In other words, unless $\partial/\partial\theta = 0$, neither the E_z nor the H_z component of the field can be identically zero. This will be discussed further in Section III.)

Assuming first a non-dissipative system, it will be seen that when H_z is zero everywhere,

$$E_z \text{ and } E_\theta \sim J_n(\lambda\rho) \begin{Bmatrix} \cos n\theta \\ \sin n\theta \end{Bmatrix}.$$

Thus the possible E-waves are determined by the boundary equation

$$J_n(\lambda a) = 0, \tag{5}$$

where

$$\lambda^2 = \gamma^2 + \omega^2/v^2.$$

This has an infinite number of real roots in λ determining an infinite number of possible waves. Only a finite number, m, of these waves will be unattenuated, however, for, if λ is to be real and γ pure imaginary, the frequency must be so high that

$$\omega/v > \lambda_{nm}, \tag{6}$$

where $\lambda_{nm}a$ is the mth root of $J_n(\lambda a) = 0$. It is therefore convenient to designate as the E_{nm}-wave that component of the E-wave for which

$$E_z \sim J_n(\lambda_{nm}\rho) \begin{Bmatrix} \cos n\theta \\ \sin n\theta \end{Bmatrix}.$$

Thus if

$$\lambda_{n,\,m+1} > \omega/v > \lambda_{nm},$$

the components $E_{n,\,m+1}, E_{n,\,m+2}, \cdots$ of the E-wave will all be attenuated but $E_{n1}, E_{n2}, \cdots, E_{nm}$ will be unattenuated. There will also be only a finite number $n + 1$ of the components $E_{01}, E_{11}, \cdots E_{n1}$, for the frequency must be at least sufficiently high so that

$$\omega/v > \lambda_{n1},$$

where $\lambda_{n1}a$ is the lowest root (excluding zero) of $J_n(\lambda a) = 0$, in order to transmit the component E_{n1} of the E-wave without attenuation. Hence the E-wave consists of a doubly terminating series of possible components; for each of the finite number $k + 1$ possible values of n there will be m_n possible values of λa or a total of

$$m_0 + m_1 + m_2 + \cdots + m_k$$

possible modes of propagation.

For the H-wave, E_z is zero everywhere,

$$E_\theta \sim J_n'(\lambda\rho) \begin{Bmatrix} \cos n\theta \\ \sin n\theta \end{Bmatrix}$$

and the possible waves are determined by the transcendental equation

$$J_n'(\lambda a) = 0, \tag{7}$$

where

$$\lambda^2 = \gamma^2 + \omega^2/v^2$$

and $J_n'(z) = (d/dz)J_n(z)$. These values of λ and consequently of γ will, of course, differ from those characterizing the E-waves. Similarly, however, there will be a doubly terminating series of possible components, H_{nm}.

Hence for both types of wave the hollow conducting cylinder constitutes a high-pass wave-filter. The critical frequency f_{nm} of the E_{nm}-wave is given by

$$f_{nm} = r_{nm}(c/2\pi a\sqrt{\epsilon\mu}), \tag{8}$$

where r_{nm} is the mth root of $J_n(\lambda a) = 0$ or $r_{nm} = \lambda_{nm}a$. Similarly for the H_{nm}-wave, the critical frequency is

$$f_{nm}' = r_{nm}'(c/2\pi a\sqrt{\epsilon\mu}), \tag{8}'$$

where

r_{nm}' is the mth root of $J_{nm}'(\lambda a) = 0$.

The propagation constant γ_{nm} is then

$$\gamma_{nm} = \frac{i\omega}{c} \cdot \frac{c}{v_{nm}'} = \frac{i\omega}{v_{nm}'}, \tag{9}$$

where the ratio c/v_{nm}' of the velocity of light in air to the phase velocity of the wave in response to any frequency f is given by

$$\begin{aligned} c/v_{nm}' &= \sqrt{\epsilon\mu}\sqrt{1 - (f_{nm}/f)^2} \\ &\to 0 \quad \text{when} \quad f \to f_{nm}, \\ &\to \sqrt{\epsilon\mu} \quad \text{when} \quad f \to \infty \end{aligned} \tag{10}$$

for the E-wave and

$$c/v_{nm}' = \sqrt{\epsilon\mu}\sqrt{1 - (f_{nm}'/f)^2}$$

for the H-wave.

For the E-wave we have

$$\begin{aligned} r_{01}, r_{02}, \cdots &= 2.405, 5.52, \cdots \\ r_{11}, r_{12}, \cdots &= 3.83, 7.02, \cdots \\ &\cdots\cdots\cdots\cdots \end{aligned}$$

and for the H-wave

$$r_{01}', r_{02}', \cdots = 3.83, 7.02, \cdots$$
$$r_{11}', r_{12}', \cdots = 1.84, 5.33, \cdots.$$

Hence, it is possible to transmit a fundamental E-wave if the radius, dielectric constant, permeability and frequency are so related that,

$$fa\sqrt{\epsilon\mu} \geqq 2.405(c/2\pi), \tag{11}$$

a fundamental H-wave provided,

$$fa\sqrt{\epsilon\mu} \geqq 3.83(c/2\pi), \tag{12}$$

the component E_{11} of the E-wave provided

$$fa\sqrt{\epsilon\mu} \geqq 3.83(c/2\pi) \tag{13}$$

and the component H_{11} of the H-wave provided

$$fa\sqrt{\epsilon\mu} \geqq 1.84(c/2\pi). \tag{14}$$

Thus from the standpoint of minimum physical constants and dimensions the component H_{11} of the H-wave is most advantageous. The consideration of the attenuation characteristics below will show, however, that this advantage is outweighed, since in practice the attenuation will be the controlling factor.

We shall now consider the characteristic impedance of the system.[12] While the derivation of the characteristic impedance is interesting and valuable on its own merits, it also provides the basis for a quasi-synthetic and approximate method of deriving the attenuation which will be developed below. The results obtained here on the assumption of a perfect conductor will be valid in the dissipative case of the next section provided the conductivity is sufficiently high so that the relation, $4\pi\sigma \gg \epsilon\omega/c^2$, obtains among the constants of the sheath.

The characteristic impedance, K, is here defined as the transverse Complex Poynting Vector, P, integrated over the cross section of the system divided by the mean square current. Thus we have, in general,

$$\begin{aligned} P &= \frac{1}{8\pi}\int dS[E\cdot H^*], \\ &= \overline{W} + i2\omega(\overline{T} - \overline{U}), \end{aligned} \tag{15}$$

[12] See the discussion of the characteristic impedance in Section I of this paper. Equation (15) below is in agreement with the definition there given.

where $\overline{W}$ is the mean energy transmitted through the cylinder, $\overline{T}$ is the mean stored magnetic energy and $\overline{U}$ the mean stored electric energy, H^* denoting the conjugate imaginary of H. Then

$$K = K_R + iK_I \tag{16}$$

and

$$\tfrac{1}{2}K_R|I|^2 = W, \tag{16a}$$

while

$$\tfrac{1}{2}K_I|I|^2 = 2\omega(\overline{T} - \overline{U}), \tag{16b}$$

I being the total current. (In a non-dissipative system $\overline{T} = \overline{U}$ and $K = K_R$.) Rewriting the integral in (15) we therefore have

$$W = \tfrac{1}{2}K_R|I|^2 = \frac{1}{8\pi}\left[\int_0^{2\pi}\int_0^a \rho(E_\rho H_\theta^* - E_\theta H_\rho^*)d\rho d\theta\right]_{\text{Real Part}}. \tag{17}$$

(From equations (1) it readily follows, that for any E- or H-wave, K may be made to depend upon either the transverse electric or transverse magnetic force alone by substituting in formula (17)

$$E_\rho H_\theta^* - E_\theta H_\rho^* = \frac{1}{c}\sqrt{\frac{\epsilon}{\mu}}\frac{v'}{v}[E]^2 = c\sqrt{\frac{\mu}{\epsilon}}\frac{v}{v'}[H]^2 \tag{18}$$

for the E-wave, and

$$E_\rho H_\theta^* - E_\theta H_\rho^* = \frac{1}{c}\sqrt{\frac{\epsilon}{\mu}}\frac{v}{v'}[E]^2 = c\sqrt{\frac{\mu}{\epsilon}}\frac{v'}{v}[H]^2 \tag{19}$$

for the H-wave, where $[E]^2$ and $[H]^2$ are defined as

$$[E]^2 = |E_\rho|^2 + |E_\theta|^2 \quad \text{and} \quad [H]^2 = |H_\rho|^2 + |H_\theta|^2.)$$

Consider first the fundamental E-wave. H_z, H_ρ and E_θ are zero and

$$\begin{aligned} E_z &= AJ_0(\rho\lambda), \\ E_\rho &= \frac{\gamma}{\lambda}AJ_1(\rho\lambda), \\ H_\theta &= \frac{\epsilon i\omega}{c^2}\frac{1}{\lambda}AJ_1(\rho\lambda), \end{aligned} \tag{20}$$

where

$$\lambda^2 = \gamma^2 + \omega^2/v^2$$

and

$$\lambda = r_{0m}/a. \quad (J_0(r_{0m}) = 0.)$$

From (3) the total axial current I_z in the sheath is given by

$$I_z = \frac{a}{2}H_\theta, \quad \rho = a. \tag{21}$$

Putting $I_z = 1$ then gives

$$A = \frac{c^2}{\epsilon i\omega} \frac{2\lambda}{aJ_1(\lambda a)}.$$

Thus

$$\begin{aligned} K &= \frac{c}{\epsilon} \frac{c}{v'} \frac{2}{a^2} \frac{1}{(J_1(\lambda a))^2} \int_0^a \rho(J_1(\lambda\rho))^2 d\rho \\ &= \frac{c}{\epsilon} \cdot \frac{c}{v'} \left[1 + \left(\frac{J_0(\lambda a)}{J_1(\lambda a)} \right)^2 - \frac{2}{\lambda a} \frac{J_0(\lambda a)}{J_1(\lambda a)} \right] \\ &= c\sqrt{\mu/\epsilon}\, \sqrt{1 - (f_{0m}/f)^2}. \end{aligned} \tag{22}$$

Now, for the fundamental component H_0 of the H-wave, E_z, E_ρ and H_θ are zero and

$$\begin{aligned} H_z &= CJ_0(\lambda\rho), \\ H_\rho &= \frac{\gamma}{\lambda} CJ_1(\lambda\rho), \\ E_\theta &= -\frac{\mu i\omega}{\lambda} CJ_1(\lambda\rho), \end{aligned} \tag{23}$$

where

$$\lambda^2 = \gamma^2 + \omega^2/v^2$$

and

$$\lambda = r_{0m}'/a. \qquad (J_0'(r_{0m}) = 0.)$$

There is no axial current transmitted by this wave but there is a circulating current in the sheath. From (4) this circulating current, I_θ, per unit length is given by

$$I_\theta = \frac{a}{2} H_z \quad \text{when} \quad \rho = a. \tag{24}$$

Thus, for the H_0-wave, we calculate the characteristic impedance with respect to unit circulating current per unit length of conductor. This gives

$$C = \frac{4\pi}{J_0(\lambda a)}$$

and

$$\begin{aligned} K &= \frac{1}{2v'} \left(\frac{4\pi\omega}{\lambda} \right)^2 \frac{1}{(J_0(\lambda a))^2} \int_0^a \rho(J_1(\lambda))^2 d\rho \\ &= (2\pi a)^2 \frac{\mu}{v'} \left(\frac{f}{r_{0m}'} \right)^2, \end{aligned} \tag{25}$$

where, as given above, r_{0m}' is the mth root of $J_0'(\lambda a)$, and, by (10),

$$v' = \frac{c}{\sqrt{\epsilon\mu}} \frac{1}{\sqrt{1 - (f_{0m}'/f)^2}}.$$

So we see that, while the characteristic impedance of the E_0-wave approaches a constant at very high frequencies, for the H_0-wave we have

$$K \sim \omega^2.$$

In other words, while the energy transmitted by the E_0-wave is independent of the frequency at sufficiently high values of frequency, that transmitted by the H_0-wave increases as the square of the frequency.

For the harmonic E- and H-waves, the currents vary as $\cos n\theta$ around the periphery of the sheath. Hence the total harmonic current is zero over any axial or normal cross-section. For these waves, however, it is possible and convenient to calculate the Complex Poynting Vector on the basis of the average mean square current intensities,

$$\frac{1}{2\pi}\int_0^{2\pi}\frac{1}{2}\left|\frac{H_\theta}{4\pi}\right|^2 d\theta \quad \text{and} \quad \frac{1}{2\pi}\int_0^{2\pi}\frac{1}{2}\left|\frac{H_z}{4\pi}\right|^2 d\theta, \qquad \rho = a,$$

which we may assume for convenience to be of the same value, 1/2, as the mean square currents associated with the fundamental components.

On this basis we shall obtain first the characteristic impedance of any harmonic component E_n of the E-wave, ignoring dissipation. Putting

$$J_n(\lambda a) = 0 \quad \text{and} \quad \lambda a = r_{nm},$$

the Complex Poynting Vector becomes

$$\overline{W} = \frac{a^4}{16c}\sqrt{\frac{\epsilon}{\mu}}\frac{v'}{v}\left(\frac{\omega}{v'}\right)^2 \frac{|A_n|^2 + |B_n|^2}{{r_{nm}}^2}(J_{n-1}(r_{nm}))^2. \tag{26}$$

On the basis of the current value which we are assuming

$$\frac{|A_n|^2 + |B_n|^2}{\lambda^2}(J_{n-1}(r_{nm}))^2 = 32\pi^2\left(\frac{c^2}{\epsilon\omega}\right)^2. \tag{27}$$

Thus

$$K_R = (2\pi a)^2 c\sqrt{\mu/\epsilon}\sqrt{1 - (f_{nm}/f)^2}. \tag{28}$$

Similarly, for the component H_n of the H-wave, we put

$$J_n'(\lambda a) = 0 \quad \text{and} \quad \lambda a = r_{nm}',$$

getting

$$\overline{W} = \frac{a^4 c}{16}\sqrt{\frac{\mu}{\epsilon}}\frac{v'}{v}\left(\frac{\omega}{v'}\right)^2 (|C_n|^2 + |D_n|^2)\frac{(J_n(r_{nm}'))^2}{(r_{nm}')^2}\left(1 - \frac{n^2}{(r_{nm}')^2}\right), \tag{29}$$

where

$$(|C_n|^2 + |D_n|^2)(J_n(r_{nm}'))^2 = 32\pi^2. \tag{30}$$

Thus

$$K_R = (2\pi a)^4 \left(1 - \frac{n^2}{(r_{nm}')^2}\right) \frac{\mu}{v'} \left(\frac{f}{r_{nm}'}\right)^2, \tag{31}$$

where, as given above, r_{nm}' is the mth root of $J_n'(\lambda a)$ and, by (10)

$$v' = \frac{c}{\sqrt{\epsilon\mu}} \frac{1}{\sqrt{1 - (f_{nm}'/f)^2}}.$$

Thus the mean transmitted energy and the characteristic impedance of all components of the H-wave increase as the square of the frequency whereas these characteristics of the E-wave are constant with respect to frequency. To appreciate the bearing of this difference upon the comparative attenuations consider the following argument.

Since the wave varies along the z-axis of the transmission system as exp. $((-\alpha - i\beta)z)$, α and β denoting the attenuation and phase constants per unit length, respectively,

$$\frac{\partial \overline{W}}{\partial z} = -2\alpha \overline{W}. \tag{32}$$

But, denoting by Q the dissipation loss per unit length of the transmission system, we have also

$$\frac{\partial \overline{W}}{\partial z} = -\overline{Q}. \tag{33}$$

Hence,

$$\alpha = \overline{Q}/2\overline{W} \tag{34}$$

$$= (4\pi \overline{Q} \int dS[E \cdot H^*]_z)_{\text{Real Part}}. \tag{35}$$

Thus, we see that, if the mean dissipation loss, $\overline{Q}$, is known or readily obtainable, the Complex Poynting Vector, $\overline{W}$, leads immediately to the attenuation.

To obtain $\overline{Q}$ we have the formula

$$\overline{Q} = -\left(\frac{1}{8\pi} \int dS[E \cdot H^*]_r\right)_{\text{Real Part}}. \tag{36}$$

Thus α may also be written

$$\alpha = \left(\frac{-\int dS[E \cdot H^*]_r}{2 \int dS[E \cdot H^*]_z}\right)_{\text{Real Part}} \tag{37}$$

in which it is evident the current is not explicitly involved. If we write

$$\overline{Q} = R(I^2)_m$$

and

$$\overline{W} = K_R(I^2)_m,$$

R being the resistance per unit length and K_R the characteristic impedance with respect to the mean square current $(I^2)_m$ we have in addition

$$\alpha = R/2K_R. \tag{38}$$

Before continuing our discussion of attenuation, we shall, therefore, have to calculate the losses in the sheath and the internal dielectric medium.

III. Dissipative Hollow Conducting Guides

In the ideal case of the preceding section, where the conductivities σ_1 and σ_2 of the dielectric and conductor are, respectively, zero and infinity, the boundary conditions are simply that $E_z = E_\theta = 0$ at the surface, $\rho = a$. When we take into account the dissipation which is actually present in the conductor (and the dielectric as well) the boundary conditions are the continuity of both the tangential electric and tangential magnetic forces. This double set of boundary conditions makes the problem inherently more difficult, of course. As we are assuming a good conductor and dielectric, we shall treat the dissipative case as a departure of the first order from the ideal case. Thus, since the dissipation has a negligible first order effect upon the phase velocity, the propagation constant γ will now be

$$\gamma = i\omega/v' + \alpha,$$

where α denotes the attenuation.

We must now consider the field in the sheath as well as the field in the inner dielectric medium. When necessary we distinguish between the electrical constants of the two media by the subscripts 2 and 1, respectively. We suppose that the sheath is electrically very thick, a legitimate assumption at the very high frequencies in which we are interested, and write for $\rho > a$,

$$\begin{aligned} E_z &= \sum_{n=0}^{\infty} K_n(\rho\lambda_2)(A_n' \cos n\theta + B_n' \sin n\theta) \exp.\ (i\omega t \pm \gamma z), \\ H_z &= \sum_{n=0}^{\infty} K_n(\rho\lambda_2)(C_n' \cos n\theta + D_n' \sin n\theta) \exp.\ (i\omega t \pm \gamma z), \end{aligned} \tag{39}$$

where

$$\lambda_2^2 = \gamma^2 - h_2^2$$

and K_n is the Bessel function of the second kind [13] (or the external Bessel function) and obtain H_ρ, H_θ, E_ρ and E_θ from (39) and (1). Putting

$$\lambda_1 a = y \quad \text{and} \quad \lambda_2 a = x,$$

and equating the tangential electric and magnetic forces E_z, E_θ and H_z, H_θ at the boundary surface $\rho = a$, we obtain eight homogeneous equations in the eight arbitrary constants. A non-trivial solution requires the vanishing of the determinant; this condition leads to the transcendental equation:

$$\left(\frac{h_1^2}{\mu_1}\frac{J_n'(y)}{yJ_n(y)} - \frac{h_2^2}{\mu_2}\frac{K_n'(x)}{xK_n(x)}\right)\left(\mu_1\frac{J_n'(y)}{yJ_n(y)} - \mu_2\frac{K_n'(x)}{xK_n(x)}\right) - n^2\gamma^2\left(\frac{1}{y^2} - \frac{1}{x^2}\right)^2 = 0, \quad (40)$$

where

$$y^2 = a^2(\gamma^2 - h_1^2) \quad (40a)$$

and

$$x^2 = a^2(\gamma^2 - h_2^2). \quad (40b)$$

The propagation constant γ is then determined by equation (40).

We mentioned in Section II that the E- or H-waves cannot exist alone in the dissipative case unless they are circularly symmetrical and it may be noticed that both E_z and H_z were required in the analysis of the preceding paragraph. To show that E_z and H_z must coexist when the conductor is dissipative, assume for the moment that $E_z = 0$. The boundary equations when $n \neq 0$ are then

$$\begin{aligned} C_nJ_n(y) &= C_n'K_n(x), & D_nJ_n(y) &= D_n'K_n(x), \\ \frac{C_n}{y^2}J_n(y) &= \frac{C_n'}{x^2}K_n(x), & \frac{D_n}{y^2}J_n(y) &= \frac{D_n'}{x^2}K_n(x), \\ \frac{\mu_1 C_n}{y}J_n'(y) &= \frac{\mu_2 C_n'}{x}K_n'(x), & \frac{\mu_1 D_n}{y}J_n'(y) &= \frac{\mu_2 D_n'}{x}K_n'(x), \end{aligned} \quad (41)$$

six equations which cannot be satisfied by four arbitrary constants. When $n = 0$, however, H_θ is everywhere zero and the boundary equations are simply

$$\begin{aligned} C_0J_0(y) &= C_0'K_0(x), \\ \frac{\mu_1 C_0}{y}J_0'(y) &= \frac{\mu_2 C_0'}{x}K_0'(x). \end{aligned} \quad (42)$$

[13] This is the Hankel function given in Jahnke und Emde, "Funktionentafeln," p. 94, 1st ed., and denoted by $H_n^{(1)}(z)$ when $\arg z < \pi$. To avoid confusion with the nth harmonic of the H-wave, we shall use K_n as a generic symbol to denote the external Bessel function.

Similarly the boundary equations can be satisfied when $H_z = 0$ provided $n = 0$ but not when $n \neq 0$.

Although E_z and H_z must co-exist in the dissipative case, one or the other will predominate in the actual wave provided the conductivity is so high that $4\pi\sigma_2 \gg \epsilon_2\omega/c^2$, a condition which is true of a good conductor unless $f \to \infty$. That this is so or, in other words, that the actual wave approximates either an E- or an H-wave will now be shown from equation (40). Since it is assumed that the conductivity is high or that

$$4\pi\sigma_2 \gg \epsilon_2\omega/c^2 \quad \text{and} \quad h_2^2 \gg \gamma^2, \tag{43}$$

$x = a\sqrt{-4\pi\sigma_2\mu_2 i\omega}$ and the asymptotic values of $K_n(x)$ and $K_n'(x)$ are valid. Equation (40) may then be written

$$\left(\frac{h_1^2}{\mu_1}\frac{J_n'(y)}{yJ_n(y)} - \frac{h_2}{\mu_2 a}\right)\left(\mu_1\frac{J_n'(y)}{yJ_n(y)} - \frac{\mu_2}{ah_2}\right) - n^2\gamma^2\left(\frac{1}{y^2} + \frac{1}{a^2h_2^2}\right) = 0. \tag{44}$$

When $h_2 = \infty$, (44) reduces to

$$J_n'(y) = 0 \text{ provided } J_n(y) \neq 0 \tag{45}$$

and to

$$J_n(y) = 0 \text{ provided } J_n'(y) \neq 0. \tag{46}$$

Thus there are two possible solutions of (44). These are in the neighborhood of $y = r$ and of $y = r'$, where r and r', respectively, are roots of $J_n(y) = 0$ and of $J_n'(y) = 0$, the equations characterizing the E- and the H-wave, respectively. We shall, therefore, refer to E- and H-waves in the dissipative case with the understanding that the actual wave approximates either one or the other type in a cylinder of sufficiently high conductivity.

As stated above, the propagation constant γ may be determined by solving equation (40). The procedure is straightforward but is complicated by the necessity for approximations and does not easily admit of physical interpretation. We may obtain the same attenuation formulas by means of the quasi-synthetic method developed at the end of Section II.

The high-frequency attenuation of the symmetric E- and H-waves is easily derived from equation (38). Here R, the resistance per unit length of the cylinder for the E-wave at sufficiently high frequencies, is given by

$$R = \frac{\sqrt{\mu_2 f/\sigma_2}}{a}. \tag{47}$$

Introducing K of (22) for K_R, and understanding that $\epsilon = \epsilon_1$, while it is assumed that $\epsilon_2 = 1$, we have

$$\alpha = \frac{1}{2ac}\sqrt{\frac{\epsilon\mu_2 f}{\mu_1\sigma_2}}\frac{1}{\sqrt{1-(f_{0m}/f)^2}} \tag{48}$$

to a high precision at high frequencies ($f > f_{0m}$). This is, of course, the contribution of the conductor and ignores the effect of the conductivity of the dielectric.

Similarly, for the fundamental H-wave, the resistance per unit length of the cylinder at sufficiently high frequencies, from equations (1) and (39) and the relations

$$\overline{Q} = \left[\frac{a}{8\pi}\int_0^{2\pi} E_\theta H_z^* d\theta\right]_{\text{Real Part}}$$

and

$$(I^2)_m = \frac{1}{2}|I_\theta|^2 = \frac{1}{2},$$

is given by

$$R = \frac{\sqrt{\mu_2 f/\sigma_2}}{a}. \tag{49}$$

Putting K of (25) for K_R, gives, to the same precision as (48), when $f > f_{0m}'$,

$$\alpha = \frac{\sqrt{\epsilon\mu_2/\mu_1\sigma_2}(f_{0m}')^2}{2ac}\frac{f^{-3/2}}{\sqrt{1-(f_{0m}'/f)^2}}. \tag{50}$$

Formulas (48) and (50), respectively, may be written in the form

$$\alpha = \frac{\alpha_0}{\sqrt{1-(f_c/f)^2}}, \qquad f > f_c \tag{48}'$$

and

$$\alpha = \alpha_0\frac{(f_c'/f)^2}{\sqrt{1-(f_c'/f)^2}}, \qquad f > f_c' \tag{50}'$$

where

$$\alpha_0 = \frac{1}{2ac}\sqrt{\frac{\epsilon\mu_2 f}{\mu_1\sigma_2}}$$

and f_c and f_c' are the critical frequencies of the fundamental E- and H-waves, respectively, as given by (8) and (8)′.

Thus, in the neighborhood of their respective critical frequencies, the attenuations of the two waves are functionally the same; ultimately, however, while the attenuation of the fundamental E-wave *increases* as $f^{1/2}$, the attenuation of the fundamental H-wave *decreases* as $f^{-3/2}$; a remarkable property peculiar to this type of wave alone.

By extending the preceding treatment to the harmonic waves, it is found after some rather laborious analysis that for all the component E-waves,

$$\alpha = \frac{\alpha_0}{\sqrt{1 - (f_n/f)^2}}, \qquad f > f_n. \tag{51}$$

Care must be taken, of course, to choose the correct critical frequency ($f_n = f_{nm}$) for the particular component wave under consideration.

For all the H-waves (including the fundamental H-wave) it is found that

$$\alpha = \frac{\alpha_0}{\sqrt{1 - (f_n'/f)^2}} \left((f_n'/f)^2 + \frac{(n/r')^2}{1 - (n/r')^2} \right). \tag{52}$$

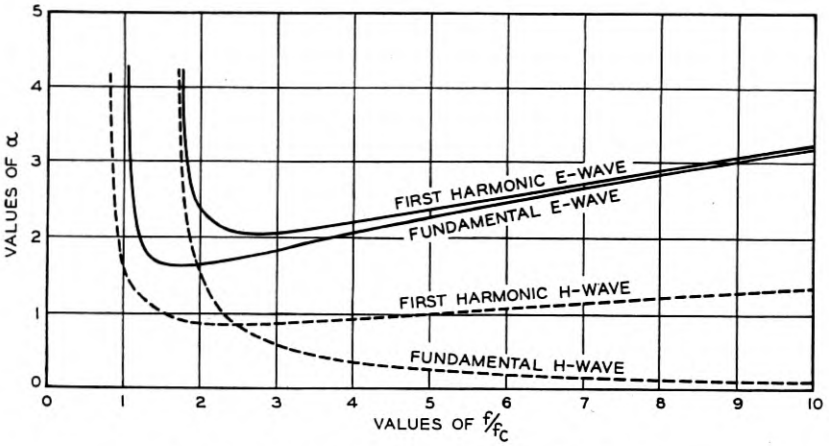
Here n is the order of the geometric harmonic wave (H_n-wave) and r' is the root of $J_n'(y)$ corresponding to the particular component wave under consideration.

The foregoing formulates the attenuation due to dissipation in the sheath alone. If we suppose that the dielectric has a very small but finite conductivity σ_1, then there must be added to the attenuation, for all types of waves, a term

$$\frac{2\pi\sigma_1 c\sqrt{\mu_1/\epsilon}}{\sqrt{1 - (f_n/f)^2}}. \tag{53}$$

To a first order approximation the dissipation has no effect on the phase velocity, which is simply v'.

Comparative values of attenuation are shown on the accompanying drawing for the fundamental and for the first harmonic E- and H-waves. This is the attenuation due to the loss in the conductor only. That due to the dielectric loss, the term given by (53), must be added. In many instances, we cannot say how large this term will be, for the losses in many dielectrics at the high frequencies involved herein are not known with any certainty at present. Such approximate calculations as we have made, however, have shown them to be very large except in the case of air.



Attenuation, α, in Hollow Conducting Cylinder.

Multiply ordinates by $A_0 = \frac{4.66}{d}\sqrt{\frac{f_c}{\sigma \times 10^4}}$ to read db per mile.

For copper, $A_0 = \frac{1.89\sqrt{f_c}}{d}$.

Multiply abscissae by $f_c = (2.30/d)10^4$ to read frequency in megacycles. Here

f_c = critical frequency of fundamental E-wave in megacycles,
d = inner diameter of cylinder in centimeters,
σ = conductivity of cylinder in emu
= 6.06×10^{-4} for copper.

IV. Dielectric Cylindrical Guides

We shall now pass to the mathematical theory of waves in dielectric "wires" of circular cross-section, immersed in air. We assume that the dielectric is perfect. The field in such a dielectric guide, and in the air outside, can be represented by the same general expressions as in hollow tubes. Thus for the nth harmonic wave, we have

$$\begin{aligned} E_z &= A_n J_n(\lambda_1\rho)\cos n\theta, \qquad H_z = B_n J_n(\lambda_1\rho)\sin n\theta, \text{ in the guide,} \\ E_z &= C_n K_n(\lambda_2\rho)\cos n\theta, \qquad H_z = D_n K_n(\lambda_2\rho)\sin n\theta, \text{ in the air.} \end{aligned} \tag{54}$$

The exponential factor $e^{-\gamma z+i\omega t}$ is implied in these as well as in the subsequent expressions for the field intensities. Another fundamental solution is obtained by changing θ into $\theta + \pi/2n$.

The transverse components of E and H are obtainable from E_z and H_z by differentiation. For our present purposes we need only E_θ and H_θ; these are

$$
\begin{aligned}
E_\theta &= \left[A_n \frac{n\gamma}{\lambda_1{}^2\rho} J_n(\lambda_1\rho) + B_n \frac{i\omega\mu_1}{\lambda_1} J_n'(\lambda_1\rho) \right] \sin n\theta, \text{ in the guide,} \\
H_\theta &= -\left[A_n \frac{i\omega\epsilon_1}{\lambda_1 c^2} J_n'(\lambda_1\rho) + B_n \frac{n\gamma}{\lambda_1{}^2\rho} J_n(\lambda_1\rho) \right] \cos n\theta, \text{ in the guide,} \\
E_\theta &= \left[C_n \frac{n\gamma}{\lambda_2{}^2\rho} K_n(\lambda_2\rho) + D_n \frac{i\omega\mu_2}{\lambda_2} K_n'(\lambda_2\rho) \right] \sin n\theta, \text{ in the air,} \\
H_\theta &= -\left[C_n \frac{i\omega\epsilon_2}{\lambda_2 c^2} K_n'(\lambda_2\rho) + D_n \frac{n\gamma}{\lambda_2{}^2\rho} K_n(\lambda_2\rho) \right] \cos n\theta, \text{ in the air.}
\end{aligned} \tag{55}
$$

The boundary conditions require the continuity of the tangential components of E and H. Hence if a is the radius of the guide, we have

$$
\begin{aligned}
A_n J_n(\lambda_1 a) &= C_n K_n(\lambda_2 a), \quad B_n J_n(\lambda_1 a) = D_n K_n(\lambda_2 a), \qquad (56) \\
A_n \frac{n\gamma}{\lambda_1{}^2 a} J_n(\lambda_1 a) &+ B_n \frac{i\omega\mu_1}{\lambda_1} J_n'(\lambda_1 a) = C_n \frac{n\gamma}{\lambda_2{}^2 a} K_n(\lambda_2 a) + D_n \frac{i\omega\mu_2}{\lambda_2} K_n'(\lambda_2 a), \\
A_n \frac{i\omega\epsilon_1}{\lambda_1 c^2} J_n'(\lambda_1 a) &+ B_n \frac{n\gamma}{\lambda_1{}^2 a} J_n(\lambda_1 a) = C_n \frac{i\omega\epsilon_2}{\lambda_2 c^2} K_n'(\lambda_2 a) + D_n \frac{n\gamma}{\lambda_2{}^2 a} K_n(\lambda_2 a).
\end{aligned}
$$

This is a homogeneous set of linear equations in the coefficients A, B, C and D from which only the ratios of these coefficients can be determined. But there are only *three* independent ratios and *four* equations; eliminating these ratios we shall obtain the *characteristic equation* of our boundary value problem from which the propagation constant γ can be calculated in terms of the frequency, the radius of the guide and the electromagnetic constants of the guide.

If $n = 0$, the above set of equations breaks up into two independent sets connecting the pairs A, C and B, D. Hence non-trivial solutions are possible by letting $A = C = 0$ or $B = D = 0$. In one case E_z is zero everywhere and in the other H_z vanishes. Thus in the circularly symmetric case we have waves of either the E-type or H-type in the sense previously defined. But if $n \neq 0$, then E_z and H_z must be present simultaneously.

The case $n = 0$ is so much simpler than the others that we shall examine it separately. Thus the characteristic equation for an E_0-wave is

$$
\frac{\epsilon_1 J_1(\lambda_1 a)}{\lambda_1 a J_0(\lambda_1 a)} = \frac{\epsilon_2 K_1(\lambda_2 a)}{\lambda_2 a K_0(\lambda_2 a)}, \tag{57}
$$

and that for an H_0-wave is

$$
\frac{\mu_1 J_1(\lambda_1 a)}{\lambda_1 a J_0(\lambda_1 a)} = \frac{\mu_2 K_1(\lambda_2 a)}{\lambda_2 a K_0(\lambda_2 a)}. \tag{58}
$$

In addition to either of these equations, we have

$$\gamma = \sqrt{\lambda_1^2 - \mu_1\epsilon_1\omega^2/c^2} = \sqrt{\lambda_2^2 - \mu_2\epsilon_2\omega^2/c^2}, \tag{59}$$

and the condition that for truly guided waves γ and λ_2 must be pure imaginary while λ_1 is real. When λ_2 is pure imaginary, the Hankel function of the second kind will decrease almost exponentially with increasing distance from the guide if this distance is sufficiently large.

If λ_1 and λ_2 are taken from (59) and substituted in (57) and (58) we shall have equations determining γ in terms of ω. Unfortunately these equations do not admit of an explicit solution for γ. It is possible, however, to carry out the numerical calculations in the following manner. We plot the left and the right terms of (57), let us say, against their arguments; then we select a pair of values of these arguments corresponding to equal ordinates. Let us suppose that we obtain

$$(\lambda_1 a)^2 = p^2, \quad (\lambda_2 a)^2 = -q^2, \tag{60}$$

where p and q are real. Referring to section III, we have $p = y$ and $iq = x$. Substituting these in (59) and solving, we have

$$\omega = \frac{c\sqrt{p^2 + q^2}}{a\sqrt{\mu_1\epsilon_1 - \mu_2\epsilon_2}}, \quad \gamma = i\sqrt{\frac{\mu_2\epsilon_2\omega^2}{c^2} + \frac{q^2}{a^2}}. \tag{61}$$

Since μ_1 usually equals μ_2, the guided waves are possible only if the dielectric constant of the guide is higher than that of the surrounding medium.

The lowest value of q is zero; the right member of (57) is then infinite and the corresponding value of p must then be a root of

$$J_0(p_m) = 0. \tag{62}$$

Corresponding to each root we have a different mode of propagation. The lowest frequency which can be transmitted in any particular mode and the corresponding propagation constant are given by

$$\omega_m = \frac{cp_m}{a\sqrt{\mu_1\epsilon_1 - \mu_2\epsilon_2}}, \quad \gamma = \frac{i\omega\sqrt{\mu_2\epsilon_2}}{c}. \tag{63}$$

At this frequency the phase velocity of propagation is equal to that of light in air. Since λ_2 is small, the field extends to great distances outside the guide. As q increases indefinitely, the right part of (57) approaches zero and p must approach the root of $J_1(x)$ near the particular root of J_0 that we happen to be considering. Thus for large values of q, we have approximately

$$\omega = \frac{cq}{a\sqrt{\mu_1\epsilon_1 - \mu_2\epsilon_2}}, \quad \gamma = \frac{i\omega\sqrt{\mu_1\epsilon_1}}{c}. \tag{64}$$

Hence at high frequencies the propagation takes place substantially with the velocity of light appropriate to the substance of the guide. The constant λ_2 being large, the field is concentrated largely in the guide.

Returning to the general n-th harmonic wave, we set

$$\begin{aligned} A_n &= SK_n(\lambda_2 a), \quad C_n = SJ_n(\lambda_1 a), \\ B_n &= TK_n(\lambda_2 a), \quad D_n = TJ_n(\lambda_1 a). \end{aligned} \tag{65}$$

Substitute in the last two equations of (56) and eliminate S and T. Thus we obtain

$$\begin{aligned} \frac{n\gamma}{a} J_n K_n \left(\frac{1}{\lambda_1^2} - \frac{1}{\lambda_2^2}\right) S &= i\omega \left(\frac{\mu_2 J_n K_n'}{\lambda_2} - \frac{\mu_1 K_n J_n'}{\lambda_1}\right) T, \\ \frac{i\omega}{c^2} \left(\frac{\epsilon_2 J_n K_n'}{\lambda_2} - \frac{\epsilon_1 K_n J_n'}{\lambda_1}\right) S &= \frac{n\gamma}{a} J_n K_n \left(\frac{1}{\lambda_1^2} - \frac{1}{\lambda_2^2}\right) T. \end{aligned} \tag{66}$$

Subsequently

$$\begin{aligned} \frac{\epsilon_1\mu_1 J_n'^2}{p^2 J_n^2} - \frac{i(\epsilon_1\mu_2 + \mu_1\epsilon_2) J_n' K_n'}{pq J_n K_n} - \frac{\epsilon_2\mu_2 K_n'^2}{q^2 K_n^2} \\ = n^2 \left(\frac{1}{p^2} + \frac{1}{q^2}\right)\left(\frac{\epsilon_1\mu_1}{p^2} + \frac{\epsilon_2\mu_2}{q^2}\right), \end{aligned} \tag{67}$$

and finally

$$\begin{aligned} \epsilon_2\mu_2 \frac{K_{n-1}K_{n+1}}{q^2 K_n^2} - \epsilon_1\mu_1 \frac{J_{n-1}J_{n+1}}{p^2 J_n^2} - \frac{i(\epsilon_1\mu_2 + \mu_1\epsilon_2) J_n' K_n'}{pq J_n K_n} \\ = n^2 \frac{\epsilon_1\mu_1 + \epsilon_2\mu_2}{p^2 q^2}. \end{aligned} \tag{68}$$

Allowing q to approach zero, we shall obtain in the limit an equation whose roots in conjunction with (61) determine the critical frequencies.

Thus if $n > 1$, we obtain

$$(\epsilon_1\mu_2 + \mu_1\epsilon_2) \frac{pJ_{n-1}(p)}{J_n(p)} = n(\epsilon_1 - \epsilon_2)(\mu_2 - \mu_1) + \frac{\epsilon_2\mu_2}{n-1} p^2. \tag{69}$$

Since ordinarily $\mu_1 = \mu_2$, (69) becomes

$$\frac{J_{n-1}(p)}{pJ_n(p)} = \frac{\epsilon_2}{(n-1)(\epsilon_1 + \epsilon_2)}. \tag{70}$$

If the dielectric constant of the guide is very much higher than that of the surrounding air, the first few roots of (70) are very close to those of $J_{n-1}(p) = 0$. As q increases indefinitely (68) degenerates into

$$\frac{J_{n-1}(p)J_{n+1}(p)}{J_n^2(p)} = 0. \tag{71}$$

Thus in the limit the roots of (68) will be *exactly* those of $J_{n-1}(p) = 0$. In other words as q varies from 0 to ∞ the corresponding value of p as given by (68) will not change much. It might appear that the limiting values of p could be roots of $J_{n+1}(p) = 0$; this is not possible, however, because in the process of transition p would have to pass through the intermediate zero of $J_n(p)$ and no real value of q is consistent with such zeros.

The case $n = 1$ requires a special examination. After multiplying (68) by q^2 and permitting q to approach zero, we find that the first term tends to infinity while the last term becomes a constant. Since the limit of $\frac{qK_1'}{K_1}$ is finite, $J_1(p)$ must approach zero. Thus for $n = 1$, the critical frequencies are determined by the zeros of $J_1(p)$.

One interesting point may be mentioned in conclusion. If the guide were surrounded by a hypothetical medium of zero dielectric constant, equation (57) for the E_0-waves would become

$$\frac{J_1(\lambda_1 a)}{\lambda_1 a J_0(\lambda_1 a)} = 0, \quad J_1(\lambda_1 a) = 0. \tag{72}$$

Thus the critical frequencies would be given by the roots of $J_1(p) = 0$ and not by those of $J_0(p) = 0$ as is the case for *any* ratio $\frac{\epsilon_2}{\epsilon_1}$ *different from zero* no matter how small it may be. Our first impression is that this result does violence to our physical common sense which demands that the hypothetical idealized case should be an approximation to the real one when one dielectric constant is large in comparison with the other. And indeed common sense is justified if one does not adhere too closely to the exact mathematical definition of the expression "critical frequency." In the region between any particular zero of $J_0(p)$, giving the true critical frequency, and the corresponding zero [14] of $J_1(p)$, giving the "approximate" critical frequency, most of the energy travels *outside* the guide, with a velocity substantially equal to that of light in the surrounding medium. The "approximate" critical frequency marks the region of the most rapid transition from wave propagation outside the guide to that inside the guide.

[14] This zero is always larger than that of $J_0(p)$.

A Magneto-Elastic Source of Noise in Steel Telephone Wires

By W. O. PENNELL and H. P. LAWTHER

The appearance of an electromotive force at the terminals of a vibrating rod or wire of magnetic material was investigated. It was concluded from experiments somewhat more simple and direct than those employed by other investigators that the effect was due to changes in the state of circular magnetization accompanying the variations of stress. The results suggested problems for more intensive investigation and applications of possible practical value.

THERE probably are few persons who have not had the experience of standing near some telephone or telegraph line out in the open and hearing the singing of the wires resulting from the wind blowing over them. Perhaps in childhood it was a source of wonderment why these sounds could not be heard at the telephones, and later, upon learning that telephone transmission was accomplished electrically and that these vibrations were mechanical only, tolerant amusement was felt at this earlier ingenuousness. Apparently it has remained until a very recent date for the discovery unmistakably to be made that it is possible under certain conditions for the mechanical vibration of a telephone wire to generate electromotive forces of sufficient magnitude to become objectionably audible in the telephone circuit. It was in April, 1935 that Mr. G. G. Jones of the Long Lines Department of the American Telephone and Telegraph Company mentioned to one of the writers the experience his Company had had a short time before in tracing down a supposed case of inductive noise to the action of the wind upon a 1200-foot steel-wire river-crossing span near Topeka, Kansas. This particular case had been cleared promptly by the application of suitable vibration damping devices generally recommended for situations where vibration might cause trouble. Special investigations then were made of long steel-wire spans at several locations in the Southwestern Bell Telephone Company territory, and it was revealed that some slight noise derived from this source actually was present in every case—and in one particular instance, where the wind velocity and direction happened to become very favorable to the production of wire vibration during the time of the inspection, the noise arose to a serious magnitude. In none of these locations had there been previous evidence that vibration was serious. That so simple and direct a phenomenon had escaped identification at the hands of telephone workers through the years of the art's existence

seemed remarkable, and especially so in view of the fact that Bell, the inventor of the telephone, in 1879 made passing note [1] of an experimental finding that probably was due in part to this effect. Accordingly, the writers' interest was aroused to the extent that an investigation was undertaken to learn the basic cause of the observed result.

In the light of subsequent knowledge it was surprising that some keen observer had not predicted and demonstrated the effect as the natural and necessary consequence of the researches of Ewing [2] and his predecessors upon the relation between state of magnetization and state of stress of a ferromagnetic specimen. Apparently it remained for von Hippel and Stierstadt [3] first to remark the phenomenon in 1931. These men were unable to interpret the effect in simple terms, however, and their reports presented a series of premature conclusions. Von Auwers [4] alone recognized the effect as capable of complete and satisfying explanation on the basis of magneto-elasticity, but he chose a method of establishing this, the interpretation of which was quite involved. For their own satisfaction in comprehending the phenomenon the present writers were led to conduct a series of experiments of qualitative character. It was felt that knowledge and clear understanding of the effect should be of immediate interest and value to workers in the general field of electrical communication in the United States.

With the aid of an amplifier having a gain of 110 decibels and terminating in a loud speaker it was practicable to conduct the experimentation with specimens of table-top dimensions. This amplifier had an input impedance very much higher than that of any of the specimens, and the response of the speaker therefore was proportional to the voltages generated by the specimens. A stretched iron wire three feet long would yield a clear sound in the speaker when its ends were connected to the input of the amplifier while it was being mechanically stimulated by plucking or bowing, and the sound from the speaker would be closely of the same quality as that heard by direct air transmission from the vibrating specimen. With this arrangement it was possible to detect any change in the magnitude of the effect as great as two to one simply by observing the loudness of the sound.

It was verified immediately that the effect must be dependent upon the property of ferro-magnetism. Taut wires of soft iron, tempered steel, or nichrome; rods of soft steel or permalloy—all produced strong sounds in the speaker when their ends were connected to the input of the amplifier while they were stimulated to vibration by bowing, plucking, or tapping. With wires of copper or brass, or with a rod of

carbon, no sound could be heard. Those small electromotive forces which must have resulted from the motion of any of the specimens in the earth's magnetic field were totally inappreciable with the apparatus employed.

Now the appearance of a potential difference between the two ends of a wire consequent to its vibration necessarily must have implied one of two situations—either there was some external influence or some relation to its surroundings which was capable of discriminating between the two ends, or else the wire possessed inherently some property that differentiated between the two directions along its length. Accordingly, exhaustive efforts were made to learn if orientation had any influence on the phenomenon. A stretched soft iron wire about three feet long yielding a clear sound in the speaker upon being plucked was employed. First, the ends of the wire were held stationary, and the wire was plucked time and again so that its plane of vibration covered representatively the various inclinations possible for this. Then there was tried a large number of positions for the axis of the wire, spread uniformly over the complete sphere. No response to orientation could be found. This negative result meant that the phenomenon under investigation must have arisen fundamentally through some condition of polarity resident within the specimen.[5]

With the phenomenon associated so definitely with the ferromagnetic property of the substance, and attributable so certainly to some quality of polarity of the specimen, it was but natural to recall [2] the considerable changes in their states of magnetization which accompany the applications of stresses with ferro-magnetic materials. In order to have produced an electromotive force along the axis, the state of magnetization of a rod or wire would have had to change in that same sense in which magnetization would have been acquired when an electromotive force was applied, and current allowed to flow, between its two ends; i.e., magnetization in closed circular paths centered upon the axis and at right angles thereto. Having formed this reasonable hypothesis of the fundamental process, experiment then was carried along the lines of testing its validity.

Any circular magnetization of the wires and rods, since its circuit would have been along paths of low reluctance wholly within the material, should have been comparatively stable and free from disturbance by external magnetic fields of moderate intensity. The observed fact that the phenomenon was quite independent of the orientation of the specimen in the earth's magnetic field was consistent with this view. The further fact that the imposition of a strong

magnetic field along the axis substantially weakened the effect was additional confirmation, for it is well known [6] of ferro-magnetic materials that strong magnetization in a given direction reduces their susceptibility in other directions.

Of course it was inferred at once that the same stresses that brought about changes in the intensities of circular magnetization of the specimens also were causing changes in the intensities of longitudinal magnetization. A coil of insulated wire was connected to the input of the amplifier. When one of the ferro-magnetic specimens was placed along the axis of this coil so that the coil winding was approximately midway between its ends, and then was stimulated to vibration, similar sounds were heard from the speaker as with the previous arrangement. As an interesting comparison it was found with a soft steel rod specimen six feet long and one half inch in diameter that a few more than fifty turns of wire on the coil were necessary to produce a sound in the speaker of the same loudness as that obtained when the amplifier input was simply connected to the two ends of the rod. Taking account of the cross-sectional areas available to the circular and to the longitudinal magnetizations, it thus was shown that the two classes of effect were not of different orders of magnitude.

There now was prepared a specimen planned especially to emphasize the effect of circular magnetization. A soft steel tube six feet long and having an external diameter of three eighths inch and a bore diameter of one eighth inch was obtained, and midway between the two ends a small opening was cut between the outer surface and the bore. Two similar windings of insulated wire were placed, each encircling closely with four turns the longitudinal wall cross-section of the tube between one end and the mid-point. Switching arrangement was provided for connecting these two windings in series either so as to encircle the total wall cross-section in one sense, or so as to encircle the two halves in opposite senses. The tube then was placed at the axis of a six-foot long solenoid, and the entire assembly was mounted with its axis horizontal and lying in the magnetic East-West direction. Sources of direct current and of sixty cycle alternating current were available.

Demagnetization of the specimen was accomplished by passing initially strong alternating current through the solenoid and the bore windings, either successively or simultaneously, and then tapering this current off uniformly to zero value. The effectiveness of the treatment could be inferred from the following observations. The specimen was made to acquire strong residual magnetism in the longitudinal direction by passing direct current momentarily through

the solenoid winding, and its magnetized condition was verified by exploration with a magnetic compass. Then upon applying the demagnetizing cycle either to the solenoid or to the bore winding this evidence of the magnetized state would disappear.

With the halves of the bore winding in series aiding and connected to the amplifier input, tapping on the end of the demagnetized tube produced a low but distinct sound in the speaker. Upon reversing one half of the bore winding the loudness of this sound usually was slightly reduced, but occasionally was slightly increased. Following the momentary passage of direct current through the bore winding with its halves connected in either sense, the loudness of the sound from the speaker was tremendously and permanently increased. Now upon reversing one half of the bore winding the loudness of this sound always was reduced markedly—although never to so low a level as that produced by the demagnetized specimen. Also, it was noted that with the bore windings in series aiding and connected to the amplifier input and starting with the tube demagnetized, the momentary passage of direct current through the solenoid winding (thereby imposing a state of residual longitudinal magnetization upon the specimen) was followed always by a moderate but marked increase in the response to tapping.

The foregoing results established quite firmly that the effect under investigation was due largely if not wholly to variations in the intensity of circular magnetization accompanying the applications of stresses. The presence of the effect to a slight degree with a specimen which presumably was in a demagnetized state remained unexplained, since testing equipment was not available for extending the inquiry further. Several plausible explanations suggested themselves. Perhaps a specimen of ferro-magnetic material could not be demagnetized completely, or—what amounted to the same thing—perhaps the state of complete demagnetization was unstable, and was followed immediately and spontaneously by the appearance of some magnetization. Again, it might have been that the state of complete demagnetization was reasonably stable of itself, but was readily disturbed by the initial application of the mechanical stresses. It seemed reasonable to expect that any such self-magnetization would have arisen most pronouncedly along the paths of least reluctance. For the time being, it was necessary to leave this point to conjecture. Certain of the results described in the paragraph immediately preceding clearly were attributable to lack of homogeneity of the specimen.

That a simple length of iron rod should be capable of functioning as a complete alternating current generator appealed to the writers

as being novel and curious. So direct a means of converting mechanical into electrical energy should find some useful applications. As the sensitive element in a telephone transmitter it should be added to the considerable list of other devices which have been used for this and allied purposes. The following arrangement was constructed. A fine iron wire was laced back and forth between pegs located along the opposite edges of a five-inch square opening in a wooden frame so as to screen the aperture with one hundred spans of the wire all in series, evenly spaced, and parallel. A sheet of paper then was cemented to the screen and the two ends of the wire were connected to the amplifier input. This device performed as a crude telephone transmitter. It was recognized, of course, that with this simple arrangement the iron wire would undergo two complete cycles of stress for each complete cycle of the air pressure upon the diaphragm, and that mechanical bias or some equally effective means would have been necessary to eliminate this distortion. Where the vibrating element itself was of magnetic material, there was the possibility of the sound source constituting its own transmitter. For example, when the amplifier input was connected between the bridge and key-head of a steel-stringed guitar the music of this instrument was reproduced quite faithfully from the speaker. This was suggestive of the possible use of the effect in studying the vibrations and strains occurring in steel structures such as bridges.

References

1. "The Bell Telephone," American Bell Telephone Company, Boston, 1908, p. 54. The following quotation from Bell's testimony was called to the writers' attention by Mr. R. I. Caughey. *Yes. I tried such an experiment, I think in the year 1879. A continuous current from a voltaic battery was passed through a stretched wire—I think a thin steel wire—which was placed in the same circuit with an ordinary commercial hand telephone. When the wire was plucked by hand, it vibrated, producing a musical tone. The hand telephone was in another room, and I listened at the telephone while a young man was plucking the wire. I heard a musical tone from the telephone at each pluck, and could recognize, also, that the character or "timbre" of the sound produced by the vibrating wire was reproduced by the telephone at my ear.*
2. Ewing, "Magnetic Induction in Iron and Other Metals," 3d ed., 1900, Chap. IX.
3. Von Hippel and Stierstadt, *Zeitschrift fur Physik*, 69, 52 (1931).
   Von Hippel, Stierstadt, and von Auwers, *Zeitschrift fur Physik*, 72, 266 (1931).
4. Von Auwers, *Zeitschrift fur Physik*, 78, 230 (1932).
5. From the results at this preliminary stage the present writers were convinced that the electromotive forces appearing at the terminals of the vibrating wire were determined by the cycles of strain, and not by the cycles of displacement, of the specimen. It seemed to them that this fact, and this fact alone, was demonstrated by von Auwers' elaborate study of the frequency and phase relationships between the mechanical vibration and the terminal electromotive forces.
6. Ewing, Chap. IX, p. 234.

An Extension of Operational Calculus

By JOHN R. CARSON

THE Heaviside operational calculus postulates at the outset that the initial (boundary) conditions at reference time $t = 0$ are those of equilibrium; that is to say, the system is at rest when suddenly energized at time $t = 0$ by a "unit" impressed force. By unit impressed force is to be understood a force which is zero before, unity after, time $t = 0$.

In a paper published in Volume 7, 1929, of the *Philosophical Magazine*, Van der Pol briefly indicated the appropriate procedure for extending the operational calculus to cover arbitrary initial conditions. The present paper is an exposition of this generalization for a system of a finite number of degrees of freedom, followed by an application to the differential equations of the transmission line. While stated in the language of electric circuit theory, it is to be understood that the processes are generally applicable to a wide variety of problems.

We start with the canonical equations for a network of n degrees of freedom

$$\begin{aligned} z_{11}I_1 + z_{12}I_2 + \cdots + z_{1n}I_n &= E_1 \\ \cdots\cdots\cdots\cdots\cdots\cdots & \\ z_{n1}I_1 + z_{n2}I_2 + \cdots + z_{nn}I_n &= E_n \end{aligned} \tag{1}$$

where

$$z_{jk} = \left(L_{jk}\frac{d}{dt} + R_{jk} + \frac{1}{C_{jk}}\int_{-\infty}^{t} dt \right). \tag{2}$$

Now multiply the equations (1) by e^{-pt} throughout and integrate from 0 to infinity. Also let J_m and F_m denote the Laplace transforms of I_m and E_m; thus

$$\begin{aligned} J_m &= \int_0^\infty I_m e^{-pt}\, dt, \\ F_m &= \int_0^\infty E_m e^{-pt}\, dt. \end{aligned} \tag{3}$$

Now let $I_m{}^0$ and $Q_m{}^0$ denote the initial values (at time $t = 0$) of I_m and the charge Q_m in the mth mesh; also let us replace z_{jk} of (2) by

$$z_{jk} = pL_{jk} + R_{jk} + 1/pC_{jk}. \tag{4}$$

We then have, replacing (1), the system of *algebraic* equations:

$$\begin{aligned} z_{11}J_1 + z_{12}J_2 + \cdots + z_{1n}J_n &= F_1 + G_1, \\ \cdots\cdots\cdots &\cdots, \\ z_{n1}J_1 + z_{n2}J_2 + \cdots + z_{nn}J_n &= F_n + G_n. \end{aligned} \tag{5}$$

In these equations, $G_1, G_2, \cdots, G_n$ denote the following summations:

$$\begin{aligned} G_1 &= L_{11}I_1^0 - Q_1^0/C_{11}p + L_{12}I_2^0 - Q_2^0/C_{12}p + \cdots \\ &\qquad\qquad + L_{1n}I_n^0 - Q_n^0/C_{1n}p, \\ \cdots&\cdots\cdots\cdots, \\ G_n &= L_{n1}I_1^0 - Q_1^0/C_{n1}p + \cdots + L_{nn}I_n^0 - Q_n^0/C_{nn}p. \end{aligned} \tag{6}$$

The right hand sides of equations (5) are thus known in terms of the impressed forces and the specified initial values of the currents and charges. They can therefore be solved in the usual manner for $J_1, \cdots, J_n$. Thus

$$J_m = \frac{F_1 + G_1}{Z_{m1}} + \frac{F_2 + G_2}{Z_{m2}} + \cdots + \frac{F_n + G_n}{Z_{mn}}. \tag{7}$$

Having thus determined $J_1, \cdots, J_n$ as functions of p, $I_1, \cdots, I_n$ are determined as functions of time by the Laplace integral equation:

$$J_m(p) = \int_0^\infty I_m(t)e^{-pt}\,dt, \qquad p_R > c, \tag{8}$$

which completes the formal solution of the problem. Note that if $G_1 = G_2 = \cdots = G_n = 0$, the solution reduces to the usual form.

Equations (6) for $G_1, \cdots, G_n$, may be written in a compact and elegant form as follows: Let

$$\begin{aligned} T &= \tfrac{1}{2}\sum\sum L_{jk}I_jI_k, \\ U &= \tfrac{1}{2}\sum\sum \frac{1}{C_{jk}} Q_jQ_k. \end{aligned} \tag{9}$$

T is then the kinetic or magnetic energy stored in the network and U is the corresponding potential or electric energy. Then

$$G_m = \left(\frac{\partial T}{\partial I_m}\right)_{t=0} - \frac{1}{p}\left(\frac{\partial U}{\partial Q_m}\right)_{t=0}, \quad m = 1, 2, \cdots, n. \tag{10}$$

The foregoing solution is compact, elegant and formally complete. In practical applications to networks of many degrees of freedom it may well present formidable difficulties in computation and interpre-

tation. This, however, is merely answerable to the complexity of the physical problem, and no simpler general solution can possibly exist.

The foregoing method when applied to the differential equations of the transmission line, leads to the following differential equations

$$\begin{aligned}(Lp + R)J &= -\frac{\partial}{\partial x}\Phi + LI^0,\\ (Cp + G)\Phi &= -\frac{\partial}{\partial x}J + CV^0.\end{aligned} \tag{11}$$

Here J and Φ are Laplace transforms of the current I and voltage V and I^0, V^0 are the initial values of I and V at reference time $t = 0$. J, Φ, I^0, V^0 are functions of x but of course independent of t.

The formal solution of equations (11) is as follows: write

$$\begin{aligned}Lp + R &= Z(p) = Z,\\ Cp + G &= Y(p) = Y,\\ \sqrt{ZY} &= \gamma, \qquad \sqrt{Z/Y} = K.\end{aligned} \tag{12}$$

Also

$$LI^0 - \frac{C}{Y}\frac{\partial}{\partial x}V^0 = F(x) = F.$$

Then

$$\begin{aligned}J = e^{-\gamma x}&\left\{A + \frac{1}{2K}\int^x dy F(y)e^{\gamma y}\right\}\\ - e^{\gamma x}&\left\{B + \frac{1}{2K}\int^x dy F(y)e^{-\gamma y}\right\},\end{aligned} \tag{13}$$

$$\Phi = -\frac{K}{\gamma}\frac{\partial}{\partial x}J + \frac{C}{Y}V^0. \tag{14}$$

A and B are constants of integration determined by the relations between J and Φ at the physical terminals of the line.

Determination of the Corrosion Behavior of Painted Iron and the Inhibitive Action of Paints *

By R. M. BURNS and H. E. HARING

THE value of paints for the protection of metal surfaces depends upon their effectiveness as physical barriers against the corrosive elements of the surrounding environment and upon the electrochemical activity of the primer pigments in rendering the surfaces passive. Physical testing methods have been developed which furnish valuable information concerning the quality and rate of aging of paint films.[1] There is, however, an obvious need for direct methods of determining the condition and behavior of the metal surface beneath the paint film, the rate of penetration of corrosive agents through the film, and the mechanism of the inhibitive action afforded by the film. The present paper describes an electrochemical method of obtaining this information.

It is well established that the process of corrosion in the presence of moisture is electrolytic in character—that it occurs by means of the operation of small galvanic cells at the surface of the metal. It should be possible, therefore, to determine the corrosion behavior of a metal by measuring the electrical characteristics of these individual corrosion cells; but the multiplicity and minute size of the anodic and cathodic areas makes such measurements impracticable. However, it is readily possible to determine the resultant of all of the polarized potentials of the anodic and cathodic areas on the metal surface, and to follow the change in this potential (of the electrode as a whole) with time.[2] Experience in this and other laboratories has demonstrated that time-potential curves obtained in this manner indicate the corrosion behavior of a metal and the state of its surface.[3] In general, it has been found that if the potential of a metal becomes more electropositive (more noble) with time, the formation of a protective film and a retardation or cessation of corrosion is taking place, while conversely, if the potential becomes more negative, solution of a protective film and acceleration of corrosion is indicated.

* Digest of a paper presented before the Spring Meeting of the Electrochemical Society at Cincinnati, Ohio, April 23–25, 1936, and published subsequently in volume 69 of its *Transactions*.

[1] Schuh, *Ind. Eng. Chem.*, **23**, 1346 (1931).

[2] Burns, *Bell System Tech. Jour.*, **15**, 20 (1936).

[3] May, *Jour. Inst. Met.*, **40**, 141 (1928); Bannister and Evans, *Jour. Chem. Soc.* (June 1930).

The corrosion of iron can be predicted by proper interpretation of such time-potential curves. The fact that the iron is painted should not alter this conclusion, and therefore the method has been applied to a study of painted iron and primer pigments.

The customary procedure for determining time-potential curves was followed with the exception that the potentiometer ordinarily employed was replaced by a vacuum tube electrometer when measurements were made of the potential of painted iron or of iron in other media of high electrical resistance. This instrument, slightly modified to take advantage of certain improvements which have been made in compensated single tube circuits, is described elsewhere.[4]

Common usage has defined iron which is corroding as "active" and iron which is not corroding as "passive." In order to obtain a background of information which might serve as a guide in the study of painted iron and paint pigments, a series of time-potential curves depicting iron and steel in the active and passive states was determined. The results are shown in Fig. 1.

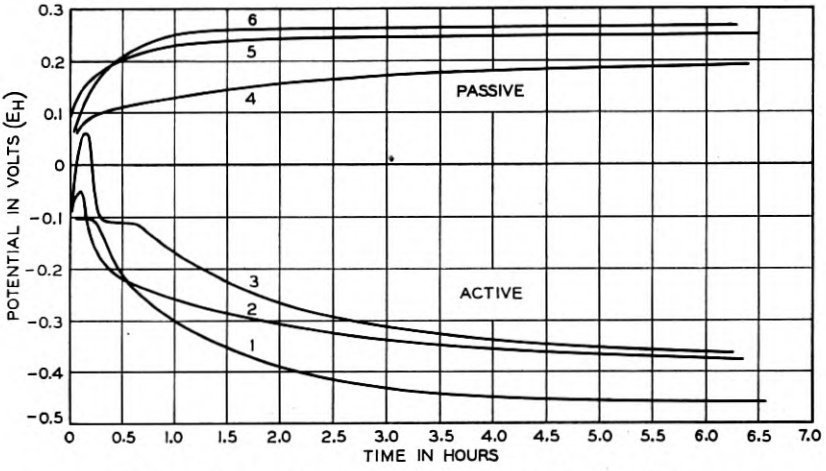


Fig. 1—Time-potential curves for active and passive iron.

Active (Corrosion)

1. Iron in tap-water.
2. Iron in 0.01N NaCl.
3. Uncleaned iron in tap-water.

Passive (No Corrosion)

4. Stainless steel in tap-water.
5. Iron in 0.01N $K_2Cr_2O_7$.
6. Uncleaned stainless steel in tap-water.

[4] Compton and Haring, *Trans. Electrochem. Soc.*, **62**, 345 (1932); D. B. Penick, *Rev. Sci. Inst.*, **6**, 115 (1935) and *Bell Laboratories Record*, **14**, 74 (1935).

It will be noted that the potentials of the test electrodes are initially quite similar, but diverge with time and form two distinct groups of curves, which ultimately become separated by about 0.7 volt. It was observed that invariably electrodes did not corrode if their potentials became more electropositive (more noble) over a long period of time, while, on the other hand, marked corrosion accompanied a negative trend of potential. A state of equilibrium was reached ultimately by the passive electrodes between 0.25 and 0.30 volt, and by the active electrodes between -0.40 and -0.45 volt.

Red oxide and red lead paints were selected for study because practical experience indicates that they are representative of the two types of protective paint, viz., (1) those which protect merely because they serve as physical barriers, and (2) those which exert a chemical inhibiting action as well.

The test electrodes were commercial iron, of high purity, in the form of 1/8 inch rods. The pigments were technical grades of red oxide (Fe_2O_3) and red lead (Pb_3O_4) of high quality. Raw linseed oil and a lead-cobalt dryer were used in the preparation of all of the paints, which were formulated and compounded in the customary manner. Approximately 20 per cent of a flexible type varnish and 10 per cent of blown linseed oil were incorporated with raw linseed oil to form the vehicle in one of the red oxide paints.

As a rule, the primary purpose of a protective paint is to shield iron from the corrosive action of water and water vapor. Total immersion is an extreme condition, but a condition to which all such paints are frequently subjected. For this reason, and also in order to speed up possible reactions and save time, all of the potential measurements on painted iron recorded in this paper were made on submerged specimens. Similar measurements on painted iron exposed to the atmosphere are equally possible and can be made without disturbing service conditions. It is planned to extend this study to include such measurements.

The time-potential curves obtained in the study of primers are presented in Fig. 2. There are included for reference typical curves (6 and 7) for iron in the active and passive states, and a curve (curve 5) for iron coated with a dried film of linseed oil. It will be noted that the linseed oil coated electrode behaved in much the same manner as bare, active iron, except that a much longer time was required for the potential changes to take place. Several days elapsed before the potential reached the equilibrium value attained by bare iron in a few hours, and at this point rust was clearly visible.

The potential of iron painted with red oxide primer (curve 1), immediately after immersion in water, was approximately 0.33 volt.

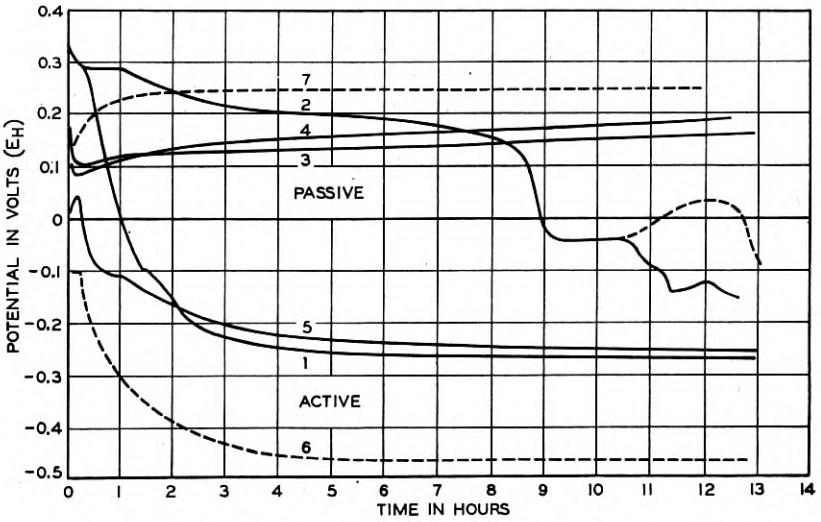


Fig. 2—Time-potential curves for painted iron in water.

Red Oxide Primer

1. Linseed oil vehicle.
2. Linseed oil plus varnish.

Red Lead Primer

3. No visible pores.
4. Visible pores.

Reference Curves

5. Dry linseed oil film.
6. Iron in tap-water.
7. Iron in 0.01N $K_2Cr_2O_7$.

In other words, the metal was passive initially and, judging from the behavior of the curve, it tended to remain so for a period of about twenty minutes. By the end of this time, water appears to have permeated the paint film in sufficient quantity to induce the active corrosion indicated by the sharp drop in potential which followed. From this time on, the curve is similar to that for the linseed oil coated electrode. Ultimate equilibrium required a much longer time, however. After several days' immersion, the paint film was removed and the iron was found to be corroded. The very gradual slope of the curves for active painted electrodes, after their sharp breaks from the passive region, may be attributed to the action of the films as partial barriers to moisture.

The impermeability of primers and their adherence to iron are known to be increased by the addition of varnish. Accordingly, red oxide

primer containing a moderate amount of varnish was subjected to test. The results are represented by curve 2 in Fig. 2.

This curve is of especial significance, not only because its characteristics are so much more pronounced than those of curve 1, but also because it furnishes an explanation for the divergence of results which have been obtained with red oxide primers in practice. The marked increase in the length of time required for this curve to pass through its various phases as compared to curve 1, is evidence of the water-excluding effect of the added varnish. The period of definite passivity has been extended to at least three times its former length, and the momentary halt in curve 1 at approximately -0.09 volt has been prolonged to an hour and a half at a slightly more positive potential. Curve 2 continues somewhat erratically in the active direction after the half-way halt in its course. Examination of the iron after several days' immersion revealed corrosion. The broken line represents, in days rather than hours, the quite different behavior of a duplicate specimen, which prior to this time had acted similarly except for the fact that it had required a somewhat longer time to pass through its various phases. Apparently a slightly less permeable paint film made it possible for the corrosive action to be stifled, temporarily at least.

Alternations of corrosive attack and film formation were observed generally when iron corroded in contact with red oxide pigments and primers. In a relatively dry atmosphere there is no doubt that iron painted with red oxide is maintained in a passive condition, but exposure to excess moisture must result eventually in active corrosion. Since, then, the protective value of red oxide primers is dependent primarily upon their ability to exclude moisture, they must be classed as physical inhibitors of corrosion.

The corrosion behavior of iron painted with red lead is clearly indicated in Fig. 2 by curves 3 and 4. The presence of a few small pores or imperfections in the paint film on one of the test electrodes did not materially affect the results. The initial potentials were somewhat lower than was the case for red oxide, and the initial trend of the curves was in the active direction, but a reversal soon took place and the iron became definitely and permanently passive. An equilibrium potential of approximately 0.25 volt was attained. Inspection of the iron after several weeks' immersion failed to reveal any sign of corrosion.

Red lead primer continues to inhibit corrosion even after moisture has fully penetrated the film. On the basis of its action, both as pigment and primer, red lead must be classed as a chemical inhibitor of corrosion. The reason for its passivating action is a disputed question. Paint chemists have inclined to the view that a highly protective lead

soap is formed, but this theory becomes untenable in view of evidence that red lead pigment alone passivates iron in much the same manner as red lead primer, and that even water solutions of the pigment have an effect. Other theories are that the iron is rendered passive by the alkalinity of the red lead, or because it is an oxidizing agent. In all probability, both of these factors are involved.

The ease with which it has been found possible to make potential measurements on painted iron with the aid of the vacuum tube electrometer, suggests the application of the time-potential method of study to the determination of the corrosion behavior of iron encased in concrete or buried underground or immersed in oil or other highly resistant media. Field study would be facilitated by substitution of a vacuum tube voltmeter for the electrometer.

The fact that there is a potential difference of at least 0.5 volt between the active and passive states of iron suggests a rapid potentiometric method for the determination of the permeability of all types of organic coatings. The time-potential curve for an iron electrode coated with the organic material and immersed in a salt solution, for example, would break sharply at the moment penetration was attained.

Abstracts of Technical Articles from Bell System Sources

The Orientation of Crystals in Silicon Iron.[1] RICHARD M. BOZORTH. X-ray examination of silicon iron prepared by N. P. Goss shows that the component crystals are oriented so that a [001] direction is parallel to the direction of rolling and a (110) plane lies in the rolling plane. This is contrary to the result reported by Goss in his paper "New Development in Electrical Strip Steels Characterized by Fine Grain Structure Approaching the Properties of a Single Crystal," published in *Transactions of the American Society for Metals*, Volume 23, June, 1935, page 511. The differences in the magnetic properties in different directions in the sheet are explained in terms of the properties of the single crystals.

Eddy Currents in Composite Laminations.[2] E. PETERSON and L. R. WRATHALL. The familiar theory of eddy current shielding leads to an expression for the impedance of a ferromagnetic core inductance coil in terms of the initial permeability and resistivity of the core material, the core geometry, and the measuring frequency. Measurements on a number of different core materials over a wide frequency range have revealed sizeable deviations from the theory in some cases. The discrepancies are especially marked in some specimens of chromium permalloy, the measured inductance over a certain frequency range being of the order of one tenth that specified by the theory.

It appears that discrepancies arise when the laminations are not homogeneous, a condition contrary to an assumption of the simple theory. The inhomogeneity takes the form of a thin surface layer which has a permeability much less than that of the interior. By etching off these surface layers, the initial permeability is increased, and discrepancies between the measured variations of impedance with frequency and those calculated for a homogeneous sheet are removed almost completely.

The theory has been extended to take account of the surface layers, and agrees well with measurements on the original unetched laminations when plausible assumptions are made regarding the properties of the surface layer.

[1] *Transactions, Amer. Soc. for Metals*, December, 1935.
[2] *Proc. I. R. E.*, February, 1936.

Applications of X-Ray Photography in Industrial Development Work.[3] J. R. TOWNSEND and L. E. ABBOTT. The fundamentals of X-ray technic as applied in developmental work of the telephone industry are outlined. A brief description is presented of the physics of X-rays, of the methods used to produce usable X-ray radiation, and of a typical industrial X-ray laboratory. Results are given of investigations at the Bell Telephone Laboratories to determine the sensitivity of X-ray methods of revealing internal defects in metals. Numerous examples illustrating the application of X-rays in telephone work are included, as well as a description of the use of gamma rays for industrial application.

Principles of Measurements of Room Acoustics.[4] E. C. WENTE. The acoustic characteristics of a room can in great part be evaluated from a knowledge of the rate with which sound in the room dies down when emission from the source ceases. The physical principles underlying the relationship are briefly discussed. It is shown by specific examples that we can obtain valuable additional information about acoustics of a room by recording the sound level at one or more points in the room when the frequency of the sound is continuously varied.

Visual Accompaniment.[5] R. WOLF. The principles of producing "Visual Accompaniments" to musical renditions for the theater are briefly described, as follows: (1) natural scenes for portraying the "musical mood" of the musical composition; (2) the changing and blending of beautiful paintings to interpret the mood, known as the Savage Method; and (3) the use of abstract color forms as a means of interpretation. The technic followed in applying the two latter methods is described in detail.

[3] Presented at the Fall 1934 mtg. of S. M. P. E., New York, N. Y.; published in somewhat condensed form in *Metal Progress*, February, 1936, under the title "Some Applications of X-Rays to Industrial Problems."

[4] *Jour. S. M. P. E.*, February, 1936.

[5] *Jour. S. M. P. E.*, February, 1936.

Contributors to this Issue

R. M. BURNS, A.B., University of Colorado, 1915; A.M., 1916; Ph.D., Princeton University, 1921; Instructor, University of Colorado, 1916–17. Second Lieutenant, Chemical Warfare Service, U. S. Army, 1918–19. Research chemist, Barrett Company, 1921–22. Western Electric Company, 1922–25. Bell Telephone Laboratories, 1925–; Assistant Chemical Director, 1931–. Dr. Burns' work has been largely in the electrochemical field and particularly on the subject of the corrosion of metals and its prevention.

JOHN R. CARSON, B.S., Princeton, 1907; E.E., 1909; M.S., 1912. American Telephone and Telegraph Company, 1914–34; Bell Telephone Laboratories, 1934–. As Transmission Theory Engineer for the American Telephone and Telegraph Company and later for the Laboratories, Mr. Carson has made substantial contributions to electric circuit and transmission theory and has published extensively on these subjects. He is now a research mathematician.

CARL J. CHRISTENSEN, B.S., Brigham Young University, 1923; M.S., University of Wisconsin, 1925; Ph.D., University of California, 1929. Instructor in Chemistry and Physics, Brigham Young University, 1923–24 and 1925–27. Bell Telephone Laboratories, 1929–. Dr. Christensen has been engaged in research in connection with microphone carbon.

H. E. CURTIS, S.B. and S.M. in Electrical Engineering, Massachusetts Institute of Technology, 1929. American Telephone and Telegraph Company, Department of Development and Research, 1929–34; Bell Telephone Laboratories, 1934–. Mr. Curtis has been engaged in high-frequency transmission problems as related to carrier systems.

KARL K. DARROW, B.S., University of Chicago, 1911; University of Paris, 1911–12; University of Berlin, 1912; Ph.D., University of Chicago, 1917. Western Electric Company, 1917–25; Bell Telephone Laboratories, 1925–. Dr. Darrow has been engaged largely in writing on various fields of physics and the allied sciences.

E. I. GREEN, A.B. Westminster College, 1915; University of Chicago, 1915–16; Member of faculty, Westminster College, 1916–17; U. S. Army, 1917–19 (Captain, Infantry); B.S., Harvard University, 1921. Department of Development and Research, American Telephone and Telegraph Company, 1921–34; Bell Telephone Laboratories, 1934–. Mr. Green has been engaged principally in work on multiplex wire transmission systems.

H. E. HARING, B.S., Franklin and Marshall College, 1916; M.A., Princeton University, 1917. Assistant Chemist, Ordnance Department, U. S. Army, 1917–19; Associate Chemist, U. S. Bureau of Standards, 1919–28; Electrochemist, Victor Talking Machine Company, 1928–29; Bell Telephone Laboratories, 1929–. Since 1919 Mr. Haring has been engaged in electrochemical research in connection with storage batteries and other electrochemical apparatus, electrodeposition, and corrosion.

H. P. LAWTHER, JR., B.A., University of Texas, 1912; M.A., Harvard, 1913; Ph.D., Harvard, 1916. Southwestern Bell Telephone Company, 1919–. Transmission and Protection Engineer, Texas Area, 1928–32; General Transmission and Protection Engineer, St. Louis, Missouri, 1932–.

F. A. LEIBE, M.E., Stevens Institute of Technology, 1922. Westinghouse Lamp Company, 1922–25; American Telephone and Telegraph Company, Department of Development and Research, 1925–34; Bell Telephone Laboratories, 1934–. Mr. Leibe has been engaged in transmission development work on carrier telephone systems.

SALLIE PERO MEAD, A.B., Barnard College, 1913; M.A., Columbia University, 1914. American Telephone and Telegraph Company, Engineering Department, 1915–19; Department of Development and Research, 1919–34. Bell Telephone Laboratories, 1934–. Mrs. Mead's work has been of a mathematical character relating to telephone transmission.

G. L. PEARSON, A.B., Willamette University, 1926; M.A., Stanford University, 1929. Bell Telephone Laboratories, 1929–. Mr. Pearson has been engaged in a study of noise in electric circuits.

W. O. PENNELL, B.Sc., Massachusetts Institute of Technology, 1896; Instructor, Lafayette College, 1896–98. Equipment and Traffic Engineer, Bell Telephone Company of Philadelphia, 1898–1902; Engineering Department, American Telephone and Telegraph

Company, 1902–03; Chief Engineer, Missouri and Kansas Telephone Company, 1903–12; Southwestern Bell Telephone Company, St. Louis, Missouri, 1912–, and for the last eighteen years its Chief Engineer. Mr. Pennell has from time to time contributed articles to various technical publications.

ARTHUR C. PETERSON, JR., B.S. in Electrical Engineering, University of Washington, 1928. American Telephone and Telegraph Company, Department of Development and Research, 1928–34; Bell Telephone Laboratories, 1934–. Mr. Peterson's work has been primarily concerned with short-wave radio transmission development.

RALPH KIMBALL POTTER, B.S., Whitman College, 1917; Artillery Corps, U. S. Army, 1917–19; E.E., Columbia University, 1923. American Telephone and Telegraph Company, Department of Development and Research, 1923–34; Bell Telephone Laboratories, 1934–. Mr. Potter has been engaged in short-wave radio transmission development work.

S. A. SCHELKUNOFF, B.A., M.A. in Mathematics, The State College of Washington, 1923; Ph.D. in Mathematics, Columbia University, 1928. Engineering Department, Western Electric Company, 1923–25. Bell Telephone Laboratories, 1925–26. Department of Mathematics, State College of Washington, 1926–29. Bell Telephone Laboratories, 1929–. Dr. Schelkunoff has been engaged in mathematical research, especially in the field of electromagnetic theory.

GEORGE C. SOUTHWORTH, B.S., Grove City College, 1914; Sc.D. (Hon.), 1931; Ph.D., Yale University, 1923. Assistant Physicist, Bureau of Standards, 1917–18; Instructor, Yale University, 1918–23. Editorial staff of *Bell System Technical Journal*, American Telephone and Telegraph Company, 1923–24; Department of Development and Research, 1924–34; Research Department, Bell Telephone Laboratories, 1934–. Dr. Southworth's work in the Bell System has been concerned chiefly with the development of short-wave radio telephony. He is the author of several papers on very high-frequency electrical phenomena.