Photo by Garber

George A. Campbell

The Bell System Technical Journal

Vol. XIV *October, 1935* *No. 4*

Dr. George A. Campbell

By F. B. JEWETT

ON the first of December next, after thirty-eight years of active and unusually productive service as a mathematical physicist and inventor, Dr. George A. Campbell retires from active membership on the staff of the Bell Telephone Laboratories.

As the history of an art can often be written most effectively in terms of the personalities who have been responsible for its upbuilding, I feel that I am not departing from the objectives of the *Bell System Technical Journal* in bringing to the attention of its readers a brief note concerning one of the chief artificers of telephone transmission. Dr. Campbell's achievements in this field entitle him beyond question to rank first among his generation of theoretical workers in electrical communication. Yet, in common with many truly great minds, it has been his nature to avoid publicity, so that outside the circle of his immediate associates and a few of the more mathematically gifted students of his chosen branch of electrical science, his fame is far from being commensurate with his achievements.

In 1897, thirty-eight years ago, the art of telephone transmission was in its infancy. Circuits of even a few hundred miles' length were rare, and the longest distance over which communication had been held was that separating New York and Chicago. It was at this time that Campbell, as a young man, after graduating from Massachusetts Institute of Technology and spending four years in graduate study at Harvard, Göttingen, Vienna and Paris, joined the staff of the American Telephone and Telegraph Company to engage in research. Familiar with the work of Rayleigh and Heaviside, Campbell's early studies sought some method of mitigating the attenuation, which levied heavy toll upon the voice currents and formed a theretofore unyielding barrier against telephone communication over very long distances.

Heaviside had suggested that inductance, if properly applied in a long telephone circuit, should diminish rather than increase the attenuation. Campbell followed this suggestion and developed a theory of loading, but in his case there occurred one of those coincidences—fortunately rare in the history of science—of two investigators arriving at substantially the same result at the same time. Independently of Professor M. I. Pupin, he worked out a complete theory of the telephone loading coil. They both applied for patents, with the result that an interference was declared, and Pupin was able to establish a slightly earlier date of conception. The loading coil interference was decided in Pupin's favor and the famous patents issued to him. The fact should be recorded, however, that Campbell's analysis of the problem—actually more detailed than Pupin's—led him to formulate rules for the design of loading coils and their spacing which were, from the very beginning, the only ones employed in this country.

As the effectiveness of telephone instruments increased and the lengths of circuits grew, noise and crosstalk became an outstanding obstacle to telephone advance. It had been shown that this crosstalk was a complex effect resulting partly from electromagnetic and partly from electrostatic induction. In unpublished memoranda written between 1903 and 1907, Campbell pointed out the importance of Maxwell's capacity coefficients in the calculation of crosstalk and coined the much-used term "direct capacity," now modernized to "direct capacitance." It was also at this time that he designed his well-known "shielded" balance, which in one form is a bridge for measuring direct capacities. He showed in these early memoranda that crosstalk between two circuits depends, to a considerable extent, and particularly in the case of loaded circuits, on a function of the various direct capacities between the wires of two circuits. He termed this function the "direct capacity unbalance." This work led to the invention of the well-known capacity unbalance test set, hundreds of which have been used in countless measurements in the manufacture and installation of toll cables.

This study of Campbell's marked an important advance since, for the precise but unwieldy theory of crosstalk, it substituted a simple approximation—an approximation which was to remain adequate until the advent of carrier systems with their higher frequencies and shorter wave-lengths.

As control was gradually extended over the characteristics of telephone circuits, both from the standpoint of their transmission effectiveness and their freedom from crosstalk and noise, the art reached the point at which development emphasis shifted to the telephone repeater. Here, an entirely new line problem arose,—namely, that of avoiding singing when repeaters are adapted to two-way amplification. Up to 1912, the only type of repeater circuit used was the so-called 21-type, in which a single repeater element amplifies messages which reach it from both directions and which requires that the two associated sections of line have very similar characteristics. A well-known limitation of the 21-type repeater is its tendency to "sing" when line unbalance or amplification exceed certain rather low limits.

On the other hand, the 22-type repeater has two amplifying elements and two artificial lines, one to balance each associated section of actual line. While the basic idea of the 22-type was old, it remained for Campbell, in a memorandum dated March 7, 1912, to reveal its properties of inherent stability. He points out that "singing will not be introduced by any possible unbalance however large, in either of the lines, provided the unbalance of the other line does not exceed a certain critical magnitude." Also his words, "the use of a compensating device such as an artificial line, to reduce the amplification at the resonant frequencies to the level of the amplification at other telephonic frequencies" suggest broadly the idea of equalizing for amplitude-frequency distortion which is brought in by the selective characteristics of the line circuits or other apparatus in a long system. Furthermore, "if it became necessary merely to eliminate certain frequencies lying outside of the range required for telephony, the use of an artificial selecting circuit" is definite anticipation of the use, subsequently common in all repeaters, of low-pass filters to cut off frequencies outside of the band transmitted and thus minimize line balance difficulties.

Campbell also indicated that the stability of a circuit as regards singing could be improved if the amplification were distributed among a number of properly spaced points along the line rather than concentrated at a single point.

Moreover, the great amplification made possible in telephone circuits by the perfection of the vacuum tube and its associated circuits permitted the use of cables for long distances with manifest advantages for congested routes. The great amplifications required for the longer cable circuits, however, could most effectively be handled by the use of "four-wire" circuits both for voice frequencies and later for carrier systems. Campbell was the originator of this type of circuit. In the same memorandum of 1912, which discussed the 22-type repeater, he suggested it as the logical extension of the one-way paths in the 22-type repeater, each path containing as many one-way amplifiers and line sections as desired.

The present situation in transmission, particularly toll transmission, is characterized by the growing use of high frequencies, the transmission of broader frequency bands and the use of the so-called "carrier" method. This we find in the most advanced form in the proposed very broad band cable circuits of the coaxial and other types. Transoceanic telephony and broadcasting are other outstanding uses. The electrical wave-filter in its many forms is one of the most important elements in all such systems. The filter appears as a means of sharply separating the currents of different continuous bands of frequencies in carrier telephone and telegraph systems, for sharp selectivity in radio systems and for various other uses in these and other forms of transmission such as telephone repeaters, telephotography, composite sets and testing apparatus. Indeed, the filter has, within the last few years, become almost as ubiquitous as the vacuum tube. The fundamental conception of the electric wave-filter arose out of Campbell's analysis of loaded lines. The patent was issued to him in 1917. It is evident to one reading his famous paper on "Loaded Lines in Telephonic Transmission," published in the *Philosophical Magazine* of 1903, that even at that time he had begun to envisage the high-pass and low-pass wave-filters.

Effective station sets are fundamental to all good transmission. In a memorandum dated October 8, 1906, Campbell disclosed the single-transformer anti-sidetone station circuit which is achieving almost world-wide acceptance. Later, he carried out a comprehensive and conclusive piece of work in revealing all of the possible circuit arrangements for doubly conjugate branches and in setting down the impedance relations of the line, network, transmitter and receiver of these various branches. This systematic analysis of the problem greatly facilitated a comprehensive survey, giving assurance that all types of circuits would be considered and that those which best fitted the available transmitters and receivers would be selected. The work was summarized in an extensive paper entitled "Maximum Output Networks for Telephone Substation and Repeater Circuits" by Campbell and Foster in the *A. I. E. E. Transactions* of 1920.

It would appear that Campbell also originated the articulation test which now finds a use wherever telephone development work is in progress. In a paper entitled "Telephonic Intelligibility," which appeared in the *Philosophical Magazine* of January, 1910, he describes how, in connection with tests he had been conducting, he made up and employed successfully articulation lists consisting of meaningless monosyllables.

This brief note is intended only to enumerate without elaboration the more outstanding of Dr. Campbell's contributions to the art of electrical communication as they fit into the history of that art. Their diversity is such as to establish the unusual versatility of Campbell's genius. His is a career unusually productive of discoveries, inventions and patents. Many of his important memoranda, however, were never worked up in the detailed form which would render them suitable for publication, and still reside only in the Company's engineering files. It would be regrettable to pass this occasion by without some notice being taken of these unpublished documents, and perhaps as fitting a commemoration as any is to print a few of the briefer ones just as they were written. Choosing somewhat at random, we are selecting the above cited memorandum of March 7, 1912, in which the 22-type repeater and the four-wire circuit are suggested, and two memoranda of earlier dates discussing capacity unbalances and crosstalk. The memorandum on repeaters is perhaps particularly interesting because of its historical flavor. Written twenty-three years ago, it refers to the measurement of attenuation in miles of cable, not decibels, and to such considerations as the natural period of the mechanical repeater diaphragm.

Dr. Campbell's Memoranda of 1907 and 1912

Introductory Note

AS mentioned in the preceding article by Dr. Jewett, the first and second of the three following memoranda were the basis of methods of designing transpositions for voice-frequency circuits. They applied particularly to non-loaded circuits but the theory was readily extended to cover loaded circuits. In an earlier and more general study written in 1904, Dr. Campbell considered the involved equations necessary to an exact solution of the crosstalk problem and deduced simplifying approximations and convenient artifices for avoiding lengthy derivations.

He first assumes a line having the circuits substantially perfectly balanced to each other by means of very frequent transpositions. He then considers the effect of an unbalanced condition in a short length of line such as might arise from an irregularity in wire or transposition spacing or an unbalanced series impedance which might be due to a poor joint. Dr. Campbell refers to such effects as "slight alterations in the impedances, mutual impedances and admittances of the system." He shows how the crosstalk can be readily computed if these alterations in impedance and admittance are known.

He then considers the case of a short untransposed length in which the coupling between circuits is systematic rather than accidental. He shows that the crosstalk in such a short length can be computed in terms of mutual impedances and admittances in just the same manner used for accidental coupling in a short length nominally perfectly balanced. He shows that the mutual impedance per unit length (which is substantially proportional to the mutual inductance) is a measure of the crosstalk effect of the magnetic field of the disturbing circuit and can be computed from a knowledge of the spacing and diameters of the wires of the disturbing and disturbed circuits. The mutual admittance per unit length is a measure of the crosstalk effect of the electric field of the disturbing circuit and is shown to be proportional to the "direct capacity unbalance" which may readily be measured or computed from measurements of the individual direct capacities. This notion of direct capacity unbalance which was deduced in the earlier memorandum of 1904 has been of the greatest usefulness in crosstalk problems both in open wire and in cable. The mutual admittance defined in this way takes account not only of the charges on the wires of the disturbing circuit but also of the charges induced by the disturbing circuit on all other wires. Thus, the electric shielding effect of other wires is taken into account. Magnetic shielding is ignored since this is unimportant with a line transposed at intervals very small compared with the wave-length. Dr. Campbell gave data for comparing the relative importance of the electric and magnetic components of the disturbing field and showed that for severe exposures both effects are of importance.

The equations given in the latter part of the memorandum of September 14, 1907, formed a basis of transposition design. They show how the crosstalk in a long transposed line may be computed with sufficient accuracy by simply summing up the effects computed individually for each short element of line. This important approximation is discussed in some detail in the earlier memorandum of 1904. Dr. Campbell prophetically says, "It must, however, always be borne in mind that we are working only with a first approximation and that in certain cases it may be necessary to continue the investigation to a higher order of approximation." In making the approximation, Dr. Campbell was, of course, thinking of voice frequency telephone circuits and at such frequencies, if the interval between transpositions is sufficiently short to guard against noise due to irregular power exposures, that interval will be but a very small fraction of the wave-length and it is unnecessary to consider the second approximation or as Dr. Campbell says to calculate "crosstalk-of-crosstalk."

When transpositions were designed for carrier frequency operation up to 30 kc. it was obviously impracticable to make a transposition interval a very small fraction of the wave-length and "crosstalk-of-crosstalk" could not be ignored. In other words, it was necessary to consider the crosstalk in each short element of line from the disturbing circuit into all the other wires on the line, the propagation of these crosstalk currents (and charges) along the line, and their effect in inducing currents in the disturbed circuit in other short elements of line. This effect has been termed interaction crosstalk since it takes account of the interaction between elements of line instead of simply summing up individual effects in each element. Thus it indeed proved true that "in certain cases" it was necessary to continue the investigation to a higher order of approximation.

I. Crosstalk Formulæ for Non-Loaded Circuits *

Take first the simple case of two perfectly symmetrical uniform circuits having the same transmission constants, which terminate at

* Memorandum dated September 14, 1907.

the same places in sets having the same impedance as the lines. Transmission upon either circuit can then give rise to no crosstalk upon the other circuit. Circuits such as two well transposed pairs on a pole lead are here to be understood; that is, there may be any number of circuits in the system but the mutual impedances and admittances between conductors connect points which are equi-spaced with respect to the impedances in each conductor.

Now suppose that at a point distanced x from the transmitting end of the circuits, slight alterations are made in the impedances, mutual impedances and admittances of the system. The effect of each change will be small and the total effect will be approximately equal to the sum of the individual results. That is, we may neglect the second-order terms, or crosstalk of crosstalk. Furthermore, unbalancing one of the given circuits alone cannot produce crosstalk. It is necessary that both circuits be unbalanced simultaneously by a single change in the system. Now, adding impedances to either side of one of the given circuits or to any third circuit will not unbalance both of the original circuits. Mutual impedance or admittance between the two sides of any circuit does not unbalance the circuit. Mutual impedance or admittance, added between either of the given circuits and any third circuit of the system, will not unbalance both of the given circuits. This leaves admittance shunted directly from one given circuit to the other given circuit and mutual impedance between the two circuits as the only source of crosstalk.

Let the admittances added between the two circuits be a, b, c, d connected between conductors 1 and 3, 3 and 2, 2 and 4, 4 and 1, respectively, where conductors 1/2 form one circuit and conductors 3/4 form the other circuit. These admittances may be resolved into the sum and difference of four admittances, as shown by the following table:

1 to 3	$a=$	$\frac{(a+b+c+d)}{4}$	$+\frac{(a+b-c-d)}{4}$	$+\frac{(a-b-c+d)}{4}$	$+\frac{(a-b+c-d)}{4}$
3 to 2	$b=$		$+$	$-$	$-$
2 to 4	$c=$		$-$	$-$	$+$
4 to 1	$d=$		$-$	$+$	$-$

By the principle of superposition the effect of the given admittances a, b, c, d will be practically the same as the sum of the effects of the four component admittances taken individually. The first component admittance $(a + b + c + d)/4$ is added symmetrically between the two wires of one circuit and the two wires of the other circuit. This will not disturb the symmetry of either circuit and will, consequently, not give rise to crosstalk. The second admittance is added between the conductor 3 and conductors 1/2 and subtracted between conductor 4 and conductors 1/2. This does not destroy the symmetry of circuit 1/2, and it can in consequence not give rise to crosstalk. The third admittance $(a - b - c + d)/4$ is added in the same way as the second with an interchange of circuits. It will also not give rise to crosstalk.

Crosstalk due to the added admittances, a, b, c, d, must therefore be due to the last component $(a - b + c - d)/4 = Y/4$ where Y is what we may call the direct admittance unbalance.

In order to determine the crosstalk occasioned by this admittance unbalance Y, when the electromotive force E is impressed upon one of the circuits, we may proceed as follows:

The circuits are connected as shown by Fig. 1. This is equivalent

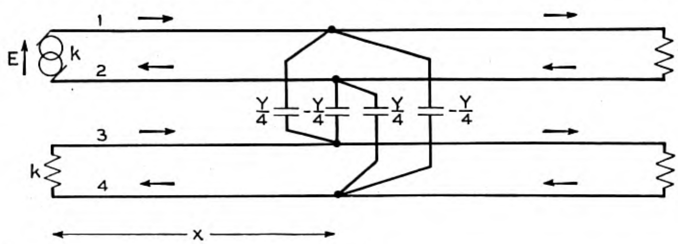


Fig. 1.

to the bridge of Fig. 2. For if the unbalancing admittances $Y/4$ were removed circuit 1/2 would be clear. Then as the e.m.f. E acts through an impedance k upon a line whose impedance is k, the potential difference at the sending end of the line would be $E/2$ and in traversing a distance x this would be attenuated by the factor $e^{-\gamma x}$. The im-

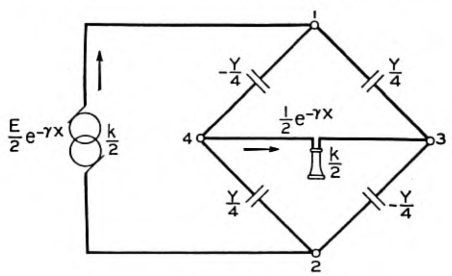


Fig. 2.

pedance of each end of circuit 1/2 at the point x will be k and therefore the entire circuit with its two ends in parallel will have the impedance $k/2$. We may therefore replace circuit 1/2 of Fig. 1 by a branch in

Fig. 2 having the impedance $k/2$ and containing an e.m.f. $Ee^{-\gamma x}/2$. Similarly circuit 3/4 will be replaced in Fig. 2 by a branch having the impedance $k/2$. As the current reaching circuit 3/4 will divide equally between the two ends of the line and the part reaching the beginning of the line will be further attenuated by the factor $e^{-\gamma x}$ the termination k in Fig. 1 is to be replaced by the receiver in Fig. 2 which indicates only $e^{-\gamma x}/2$ of the current flowing through it. Substituting these values in the expression for the galvanometer current in a bridge [1] we find for the crosstalk current

$$\Delta I_y = -\frac{EY}{16} e^{-2\gamma x} \left[\frac{1}{1 - \frac{k^2 Y^2}{64}} \right]$$

or approximately

$$\Delta I_y = -\frac{EY}{16} e^{-2\gamma x} \text{ as } Y \text{ is very small.}$$

This is the current at the end corresponding to the transmitting station. At the other end the attenuation factor will be that corresponding to transmission over the entire length of line l—making the crosstalk

$$\Delta I_y' = \frac{EY}{16} e^{-\gamma l}.$$

It will be noticed that the crosstalk at the farther end of the line is independent of the position of the admittance unbalance, while the crosstalk at the transmitting end of the system will diminish as the point of unbalance is moved farther from this end. In one case the wave must traverse the entire distance between terminal stations; in the other it must travel down the line to the point where it is carried across from one circuit to the other and then back from this point to the beginning of the line where the crosstalk is received.

The mutual impedances between the four conductors composing the two circuits may be treated in a manner similar to that which has been employed for the admittances between these conductors. Assume that any four mutual impedances are added and then divide them into four components, of which three may be shown to give rise to no crosstalk. The remaining component is the mutual impedance unbalance $Z/4$ and the crosstalk due to it may be found as follows. The circuit is shown by Fig. 3, which may be replaced by the trans-

[1] Maxwell I, § 347.

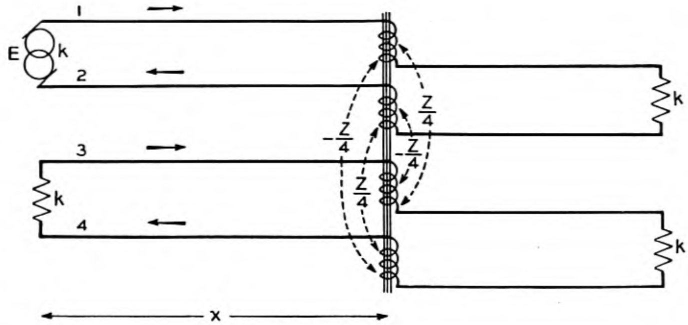


Fig. 3.

former circuit of Fig. 4. If there is no unbalance Z so that we have the original uniform system the current $E/2k$ starts out on 1/2 and at the point of mutual impedance unbalance it becomes $Ee^{-\gamma x}/2k$. Circuit 1/2 therefore behaves like a primary of impedance $2k$ (for the two ends of the line are in series) containing e.m.f. $Ee^{-\gamma x}$. The circuit

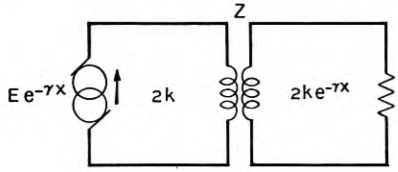


Fig. 4.

3/4 acts as a secondary of impedance $2k$. If the mutual impedance between the two lines is now made Z without any other change the current on 3/4 at x is

$$-\frac{EZe^{-\gamma x}}{4k^2 - Z^2},$$

which is attenuated by the factor $e^{-\gamma x}$ in reaching the transmitting end of the line, making the crosstalk

$$\Delta I_z = -\frac{EZ}{4k^2}e^{-2\gamma x}\left[\frac{1}{1-\dfrac{Z^2}{4k^2}}\right]$$

or approximately

$$\Delta I_z = -\frac{EZ}{4k^2}e^{-2\gamma x} \text{ as } Z \text{ is small.}$$

At the farther end of the listening circuit, the approximate expression is

$$\Delta I_z' = -\frac{EZ}{4k^2} e^{-\gamma l}.$$

The total crosstalk due to the admittance and mutual impedance unbalance Y and Z is thus:

$$\Delta I = -E\left[\frac{Y}{16} + \frac{Z}{4k^2}\right] e^{-2\gamma x} \text{ at the transmitting end,}$$

$$\Delta I' = +E\left[\frac{Y}{16} - \frac{Z}{4k^2}\right] e^{-\gamma l} \text{ at the receiving end,}$$

from which we see that the ratio of the crosstalk due to the two kinds of unbalance is independent of the location of the point on the circuit at which these unbalances are introduced, and the two will be numerically equal in case:

$$Y = \frac{4}{k^2} Z$$

or

$$Y = \frac{4}{(667)^2} Z = 9 \cdot 10^{-6} Z, \text{ on non-loaded open wires.}$$

Appended tables give the direct capacity unbalances and the mutual impedance unbalances for a 40-wire lead of No. 12 wire. The relative importance of the two in producing crosstalk is shown below for a few typical cases:

	Unbalance per Mile		Ratio $\frac{9 \cdot 10^{-6} Z}{Y}$
	Direct Capacity Unbalance mmf.	Mutual Inductance Unbalance mh.	
Pairs adjacent on same cross-arm			
1/2 and 3/4	−1014	− 93	0.8
3/4 and 5/6	−1469	−143	0.9
Pairs adjacent in vertical plane			
1/2 and 11/12	602	72	1.1
3/4 and 13/14	409	72	1.6
5/6 and 15/16	704	147	1.9
Pairs separated by two steps			
1/2 and 5/6	− 116	− 31	2.4
3/4 and 7/8	− 84	− 19	−2.0
1/2 and 13/14	− 93	2	−0.2
3/4 and 15/16	− 150	4	−0.2
1/2 and 21/22	136	20	1.3
5/6 and 25/26	46	43	8.4

It will be noted that the two are of about equal importance for the cases of most severe static exposure. For 5/6 and 15/16, which is quite a severe exposure, the magnetic is twice as important as the static. For the pairs which are so far removed as to bring in considerable static shielding from the other wires the magnetic crosstalk may be still more important relatively. Thus for 5/6 and 25/26 it is eight times as great as the static.

At the sending end of the system the static and the magnetic crosstalks combine, while at the other end they tend to cancel each other. In practice, therefore, the summation is the more important case.

If both unbalances are pure reactances, the one being pure capacity and the other pure inductance, and the line has approximately the same impedance at all frequencies, the character of the crosstalk will be the same whether it is produced by capacity or mutual inductance. This will be approximately the case on well insulated open-wire lines. On non-loaded cable circuits the line impedance decreases as the frequency rises. On cables, therefore, the crosstalk due to mutual impedance will have the higher frequencies more strongly pronounced than the crosstalk due to capacity.

For a transposed line we find the total crosstalk by integrating the crosstalk throughout the entire length of the line. We will assume that the lines are infinitely long and that the transpositions give the system a periodic structure of lengths. Let x be the distance to the first transposition, a, b, c, $d \cdots s$, the distances from the transposition to the others in the periodic section.

$$I = -E\left(\frac{Y}{16}+\frac{Z}{4k^2}\right)\int_{0,\ x+a,\ x+a,\ \ldots}^{x,\ x,\ x+b,\ x+b,\ \ldots} e^{-2\gamma x}dx$$

$$= -\frac{E}{2\gamma}\left(\frac{Y}{16}+\frac{Z}{4k^2}\right)(1 - 2e^{-2\gamma x} + 2e^{-2\gamma(x+a)} \cdots - 2e^{-2\gamma(x+s)} + \cdots),$$

since the periodic section s must contain an even number of transpositions.

$$I = -\frac{E}{2\gamma}\left(\frac{Y}{16}+\frac{Z}{4k^2}\right)\left(1 - 2e^{-2\gamma x}\frac{1 - e^{-2\gamma a} + e^{-2\gamma b} - \cdots}{1 - e^{-2\gamma s}}\right)$$

or approximately

$$I = -\frac{E}{2\gamma}\left(\frac{Y}{16}+\frac{Z}{4k^2}\right)\left\{ 1 - 2[1 - 2\gamma x + 2\gamma^2 x^2] \right.$$

$$\times \frac{2\gamma(a - b + c\cdots) - 2\gamma^2(a^2 - b^2 + \cdots) + \frac{4\gamma^3}{3}(a^3 - b^3 + \cdots)}{2\gamma s - 2\gamma^2 s^2 + \frac{4}{3}\gamma^3 s^3}\Bigg\}.$$

But the sum of alternate intervals in the transposition periodic section s must equal $s/2$. Therefore $a - b + c\cdots = a + (c - b) + (e - d) + \cdots = s/2$, and to the same approximation,

$$I = -\frac{E}{2\gamma}\left(\frac{Y}{16} + \frac{Z}{4k^2}\right)\Bigg\{1 - [1 - 2\gamma x + 2\gamma^2 x^2]$$

$$\times \frac{1 - \frac{2\gamma}{s}(a^2 - b^2 + \cdots) + \frac{4\gamma^2}{3s}(a^3 - b^3 + \cdots)}{1 - \gamma s + 2\gamma^2 s^2/3 - \cdots}\Bigg\}$$

$$= \frac{E}{2}\left(\frac{Y}{16} + \frac{Z}{4k^2}\right) s \Bigg[\left\{1 - \frac{2x}{s} - \frac{2(a^2 - b^2 + \cdots)}{s^2}\right\}$$

$$+ \left\{\frac{1}{3} - \frac{2x}{s} + \frac{2x^2}{s^2} - 2\left(1 - \frac{2x}{s}\right)\frac{a^2 - b^2 + \cdots}{s^2}\right.$$

$$\left. + \frac{4}{3s^3}(a^3 - b^3 + \cdots)\right\}\gamma s\Bigg].$$

It will be noticed that the crosstalk varies linearly with the distance from the first transposition, approximately, and that by a suitable choice for this distance the crosstalk may be reduced to zero to the first approximation. This is, however, not a matter of especial practical importance, for incidental irregularities contribute to the crosstalk in practice. As soon as the crosstalk due to the regular transposition system is reduced to the order of that due to the accidental irregularities further reduction of this crosstalk is not a matter of commercial importance. The accidental irregularities in the distribution of the wires therefore set a limit to the extent to which it is worth while to reduce the length of the transposition sections.

If there are but two transpositions in the periodic interval $a = s/2$, $b = c = \cdots = 0$

$$I = \frac{E}{2}\left(\frac{Y}{16} + \frac{Z}{4k^2}\right) s \left[\left(\frac{1}{2} - \frac{2x}{s}\right) - \frac{x}{s}\left(1 - \frac{2x}{s}\right)\gamma s\right] \text{ approximately.}$$

If $x = s/4$, I vanishes as to first order terms. If $x = s/2$ or 0, I has its maximum value,

$$\frac{E}{2}\left(\frac{Y}{16} + \frac{Z}{4k^2}\right)\frac{s}{2}.$$

II. Crosstalk Formulæ for Phantom Circuits *

There may be crosstalk between two phantom circuits, between a phantom and a distinct two-wire circuit, or between a phantom and one of its own side circuits.

In the first case we may assume that each side of both phantoms is perfectly balanced with respect to every part of the system, as we are not here concerned with the crosstalk on the side circuits. The two wires forming the side of a phantom may then be treated as a single conductor. The sources of crosstalk will therefore be direct admittance unbalance and mutual impedance unbalance exactly as for ordinary pairs. The capacity unbalance in formulæ [1] is found from the 16 direct capacities between the two sets of four wires, and the mutual impedance unbalance is found from the 16 mutual impedances between the same wires, but this is to be divided by four in order to allow for the division of the current between the two wires on each side of both phantom circuits.

In the second case each side of the phantom circuit may be assumed perfectly balanced. In the computation of the capacity unbalance, 8 direct capacities enter. There are also 8 mutual impedances involved in the mutual impedance unbalance, and these must be divided by 2 in order to allow for the division of the phantom circuit current between the two wires.

The crosstalk between a phantom circuit and one of its side circuits differs materially from the others, as the use of the same conductors to form the side circuit and one side of the phantom circuit introduces two additional sources of unbalance. These are: unbalance in the impedance of the two conductors forming the side circuit, and unbalance in the direct admittance from the two conductors forming the side circuit to the system outside of the phantom conductors.

The assumed distribution of unbalances is shown by Fig. 1. Insert

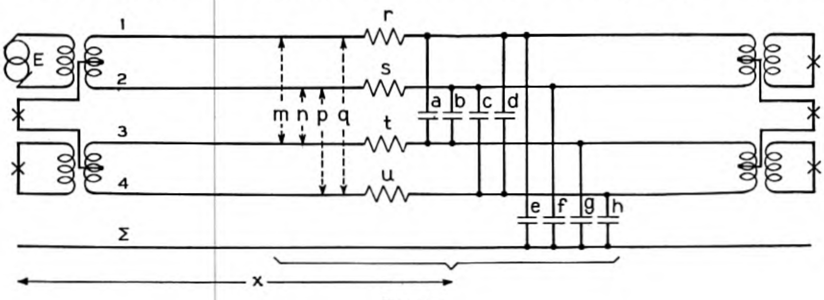


Fig. 1.

* Memorandum dated October 31, 1907.
[1] See September 14 memorandum.

an e.m.f. (E) at the sending end of circuit 1–2. Then at distance x conductor 1 will be at the potential $Ee^{-\gamma x}/4$ and carry the current $Ee^{-\gamma x}/2k$ before the unbalances are introduced. The potential of 2 and the current carried by 2 will be the same with sign reversed. Conductors 3, 4 and all others (Σ) in the system will be at potential 0 and carry no current. It follows at once that the impedances in 3 and 4 (t, u) and the admittances between these conductors and the conductors (Σ) of the system (g, h) will contribute nothing towards the crosstalk between 1–2 and the phantom. As equal impedances inserted in 1 and 2 will not unbalance the side circuit, the crosstalk must depend upon the difference between r and s and in consequence we may substitute $(r - s)/2$ for r in 1 and the negative of this in 2 without altering the crosstalk. Similarly, the effect of e and f depends entirely upon their difference, and we may substitute $\pm\ (e - f)/2$ for e and f. The direct capacities (a, b, c, d) may be resolved into four components,[2] and of these only the third unbalances both 1–2 and the phantom. The same applies to the mutual impedance unbalance. It follows that the crosstalk depends solely upon

$$\begin{aligned} X &= (r - s), \\ Y' &= (e - f), \\ Y'' &= (a - b - c + d), \\ Z &= (m - n - p + q), \end{aligned}$$

as indicated in Fig. 2.

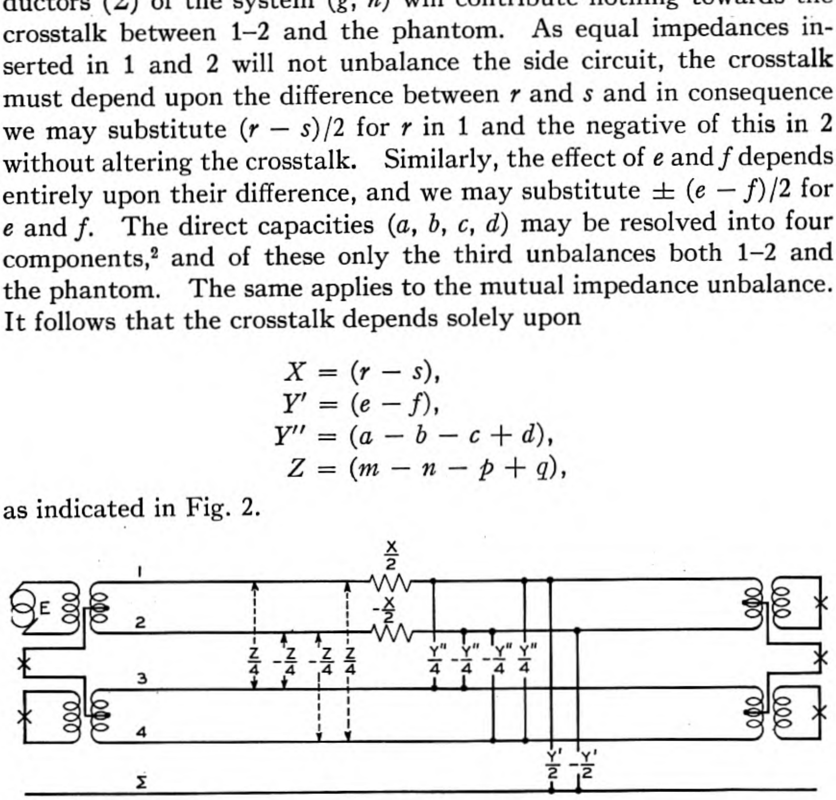


Fig. 2.

Substituting e.m.f. and currents for these in accordance with the rules for small changes, we have Fig. 3. This system is perfectly symmetrical with respect to the two wires in each side of the phantom and we may now treat the two wires on each side as one conductor, as in Fig. 4. The total e.m.f. around the phantom is $Ee^{-\gamma x}(X + Z)/4k$ and the impedance of the two ends of the phantom in series is $2K$.

[2] See September 14 memorandum.

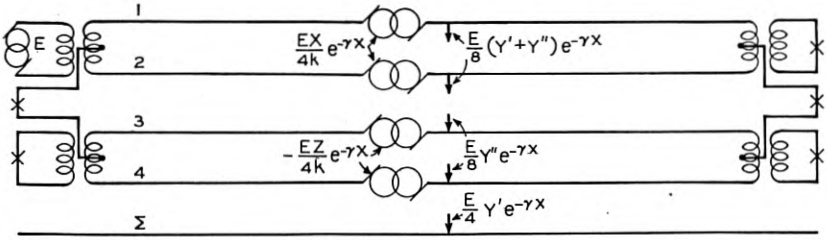


Fig. 3.

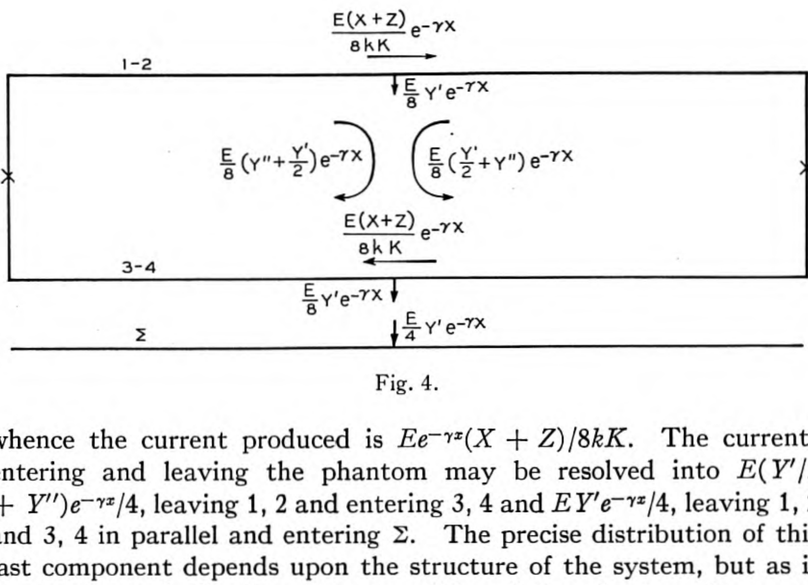


Fig. 4.

whence the current produced is $Ee^{-\gamma x}(X + Z)/8kK$. The currents entering and leaving the phantom may be resolved into $E(Y'/2 + Y'')e^{-\gamma x}/4$, leaving 1, 2 and entering 3, 4 and $EY'e^{-\gamma x}/4$, leaving 1, 2 and 3, 4 in parallel and entering Σ. The precise distribution of this last component depends upon the structure of the system, but as it flows symmetrically down the two sides of the phantom, it cannot introduce crosstalk. The first component divides equally between the two ends of the line, and this is the capacity unbalance current. The currents are therefore completely shown in Fig. 4. Allowing for the attenuation of the currents in reaching the ends of the line, we have the formulæ given below:

If a phantom (1, 2–3, 4) and its side circuit (1–2) have the transmission constants (K, Γ) and (k, γ) and the length (l) and are symmetrical throughout, then the crosstalk which will be introduced between them by the addition, at the distance (x) from the transmitting end, of small impedances (r, s, t, u) in the conductors, mutual impedances (m, n, p, q) between the conductors of the two side circuits and direct admittances $(a, b, c, d; e, f, g, h)$ between the conductors of the two side circuits and between these conductors and the remainder of the system is:

$$\Delta I = E\left(\frac{X+Z}{8kK} + \frac{Y}{16}\right) e^{-(\gamma+\Gamma)x} \quad \text{at the sending end,}$$

$$\Delta I' = E\left(\frac{X+Z}{8kK} - \frac{Y}{16}\right) e^{-[(\gamma-\Gamma)x+\Gamma l]} \text{ at the distant end,}$$

where $X = (r - s)$,

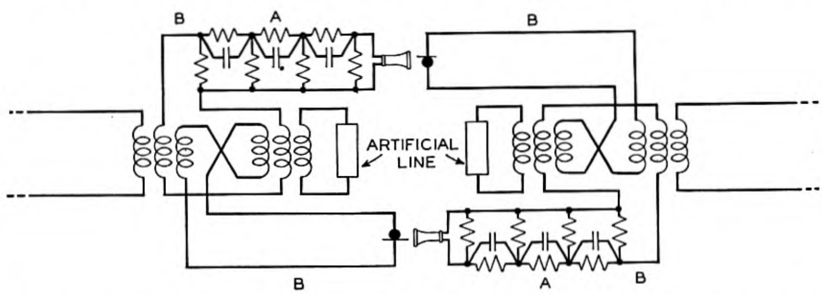
$$Y = 2(a - b - c + d) + (e - f),$$
$$Z = (m - n - p + q),$$

and the positive direction in both circuits is the same as in conductor 1.

III. Repeater Circuits *

The following points seem to merit an experimental trial.

1. A Two-way Repeater Circuit Including Two Repeaters, Each Operating as a One-way Repeater Only, as Illustrated by the Following Sketch



With this circuit the allowable unbalance is about double that with our present standard circuit. In addition to this, singing will not be introduced by any possible unbalance, however large, in either of the lines, provided the unbalance of the other line does not exceed a certain critical magnitude. Furthermore, the two lines connected together may differ radically in character since each is balanced separately against its own artificial line.

The present standard circuit or any one of several other repeater circuits may be substituted in place of the basic circuit shown in this sketch.

Although the circuit requires that all of the repeating apparatus be duplicated and that two artificial lines (of which at least one must be a close copy of the corresponding actual line as regards telephonic

* Dated March 7, 1912.

sending end impedance) be added, it still seems to me probable that the improved performance of the circuit will prove in this extra equipment.

2. The Use of Repeaters of Small Amplification at Periodic Intervals along the Line

Theoretically, a given total amplification can be secured with a larger singing margin if it is distributed among a number of properly spaced points along the line rather than concentrated at a single point. For example, four equally spaced repeaters each giving an amplification of five miles might be substituted for a single repeater giving twenty miles. If the circuit suggested by the above sketch were employed this would mean a total of eight repeater elements of which four would be used, one after another, as one-way repeaters in each direction. This raises the old question as to whether equally good quality can be obtained when several repeaters are used in securing a given amplification. This point seems worth further direct experimental investigation; one step in the right direction has probably been made by raising the natural period of the diaphragm.

3. The Use of a Compensating Device Such as an Artificial Line to Reduce the Amplification at the Resonant Frequencies to the Level of the Amplification at Other Telephonic Frequencies

In the sketch, equalizing artificial lines are shown at *AA*; obviously the same result may be secured by introducing them at any of a number of other points in the circuit. In this way the singing margin can be increased and the quality be somewhat improved, without materially reducing the telephonic amplification. But on general principles it would seem desirable to carry the equalization as far as possible in the repeater itself. The variability of the repeater sets a limit to what may be accomplished by any compensating device which reduces the total amplification by an invariable amount at each frequency and thereby increases the percentage variation. If it became necessary merely to eliminate certain frequencies lying outside of the range required for telephony, the use of an artificial selecting circuit would seem to present no difficulty.

The variation as well as the average amplification obtained from repeaters should be investigated. When these data have been obtained for the best type of repeater it will be possible to determine whether any material benefits can be derived by the introduction of compensating circuits.

4. Use of Two Lines for a Portion of the Route

The circuit shown above enables us to switch two-way transmission from a single line to a pair of lines and vice versa, since any amount of line may be inserted at *BBBB*. In connection with these lines any number of additional one-way repeaters may be inserted. The operation of the system is left unchanged beyond the change in the effective amplification which is equal to the difference between the repeater amplification and the attenuation of the inserted line. In case the total attenuation of the pair of lines exceeds the total amplification of all the repeaters at every frequency the system cannot sing whatever lines be connected at the ends. Practical applications will hinge upon the possibility of securing good quality from a number of repeaters used in sequence.

Suggestions 1 and 2 seem sufficiently promising to warrant some experimental work at an early date.

I am preparing a discussion of the general repeater circuit including any number of repeating elements and shall present the theoretical deductions applicable to the above suggestions in that memorandum.

Some Aspects of Low-Frequency Induction Between Power and Telephone Circuits *

By H. R. HUNTLEY and E. J. O'CONNELL

This article discusses the phenomena involved in low-frequency induction between power and telephone circuits and describes a demonstration which has been developed to illustrate certain of them.

Introduction

IN the practical problem of inductive coordination between power and telephone circuits, there are two aspects to be considered:

1. Induction within the frequency range used in transmitting speech which may result in disturbing noise in telephone receivers. This phenomenon is usually associated with the normal operation of power and telephone systems although abnormal conditions in either system may result in a large increase in the noise.
2. Induction at the fundamental frequencies used in the transmission of power. This is commonly referred to as "low-frequency induction" and in some cases may reach such magnitudes as to interrupt telephone service, constitute a hazard to telephone employees and produce other detrimental effects. Induced voltages of magnitudes sufficient to cause operating difficulties in telephone circuits occur usually only under abnormal power circuit conditions which produce large currents in the earth. Under normal circuit conditions, three-phase power circuits are so nearly balanced with respect to ground at their fundamental frequency that induction at this frequency is rarely sufficient to seriously affect well balanced telephone circuits.

Both types of induction have been and are being intensively studied cooperatively by the power and telephone industries. Much of this work has been handled through the Joint General Committee of the National Electric Light Association and Bell Telephone System which was formed in 1921 and it is now being carried forward by the Joint General Committee of the Edison Electric Institute and Bell Telephone System.

* This paper appeared in somewhat different form in *Amer. Railway Assoc. Proc.*, June, 1934, under the title "Demonstration of Low-Frequency Induction Between Power and Telephone Circuits" by H. R. Huntley.

Since the inductive coordination of power and telephone plants inherently involves the characteristics of both systems as well as the physical relations between them, problems can be effectively handled only by joint consideration in each specific case. As pointed out in a previous article dealing with noise induction,[1] effective cooperative action depends upon an adequate mutual understanding of the principles involved in coordination. Many of these principles can be demonstrated using comparatively simple apparatus.

This article describes a demonstration that has been developed to illustrate some of the more important factors concerned with low-frequency induction between power and telephone circuits, together with a discussion of this subject along the lines which would be followed in presenting the demonstration.

Demonstration Apparatus

The demonstration apparatus consists of two separate arrangements as follows:

(*a*) For many of the demonstrations, a miniature inductive exposure, consisting of a three-wire power line and a two-wire telephone line can be used. The power line is energized at a comparatively low voltage from a three-phase bank of transformers. The telephone line can be grounded at either or both ends and by means of a voltage measuring device, consisting of an amplifier and a projecting meter, a qualitative indication of the voltage along it or between it and ground can be obtained.

(*b*) For other demonstrations, a fairly high voltage in the telephone circuit is required, but since it is impracticable to secure this voltage using the miniature inductive exposure, it is necessary to use an iron core transformer. In order to improve the safety conditions when using this higher voltage, the miniature lines are not used and the circuit is entirely separate from that used in the low-voltage demonstrations.

A power supply frequency of 60 cycles per second is used. The phenomena illustrated are, however, applicable for all other frequencies commonly encountered in power transmission and distribution circuits.

In a demonstration of this kind, where the exposure is compressed into a small space and where the amount of power available is limited, it is obvious that the results can have no quantitative significance. This demonstration, therefore, is designed only to provide qualitative illustrations of some of the principles involved.

[1] "Some Theoretical and Practical Aspects of Noise Induction," by R. F. Davis and H. R. Huntley, published in *Bell System Technical Journal*, October, 1933.

Fundamentals of Problem—Magnetic Induction

Induction arises due to the fact that any wire transmitting electricity is surrounded by electric and magnetic fields which may cause voltages to appear on other wires in these fields. The relative strengths of the electric and magnetic fields depend on the characteristics of the circuit, the former being a function of the voltage on the circuit and the latter a function of the current in it. Induction due to electric fields is commonly called "electric induction" while that due to magnetic fields is called "magnetic induction."

When a ground occurs on a power line there are two factors which influence the induction into neighboring telephone circuits:

(*a*) The residual voltage is increased, which increases the electric induction.

(*b*) The residual current is increased, which increases the magnetic induction.

Both from theoretical analyses and experience it is known that magnetic induction is more important than electric induction in most cases of low-frequency induction. Consequently, the demonstration is concerned only with magnetic induction, i.e., induction due to the power system currents.

The magnetic field about a wire faulted to ground and carrying fault current is shown in Fig. 1. This magnetic field varies in proportion

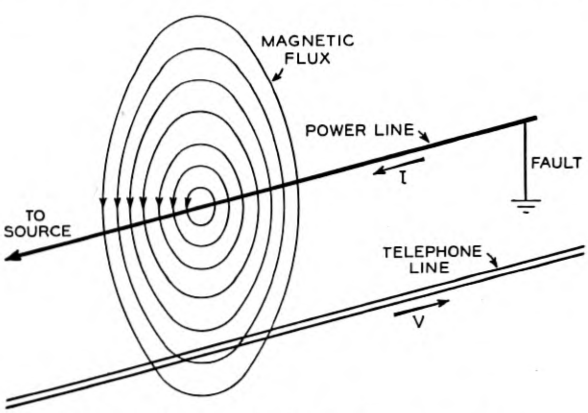


Fig. 1—Process of induction from currents.

to the current in the wire causing it. If other conductors, such as the pair of telephone wires shown in Fig. 1, lie within this field there are induced along them voltages proportional at every instant to the time rate of change of the magnetic flux which links the wires.

Consider first that the inducing circuit carries direct current. In such a case, as long as the magnitude of the direct current does not change, the magnetic flux is constant and no voltage will be induced in a paralleling telephone wire. However, if the current in the inducing wire changes suddenly, the flux about it changes in the same proportion and voltage is induced momentarily in the paralleling wire. The magnitude of this impulse of voltage will be proportional to the rate at which the flux changes and the voltage will last only as long as the flux is changing. Rapid changes in the current in direct current circuits, with consequent voltage impulses on paralleling telephone circuits, may occur when power apparatus is turned off or on. Also, of course, if a short circuit occurs, the current may rise very rapidly and then fall very rapidly as the circuit breaker operates. While the consideration of these direct current phenomena are important in some situations, they are not included in this demonstration and will not be further considered.

In the alternating current case, the current in the inducing wire is continually alternating so that the flux about it is continually alternating. Consequently, there will be induced in a paralleling wire, an alternating voltage proportional to the inducing current. It should be noted particularly that the induced voltage acts *along* the wire rather than between the wire and ground.

General Nature of Phenomena

In applying these principles of magnetic induction to the low-frequency induction problem, only the conditions which exist when a power circuit is faulted to ground and before the current is interrupted (usually by the operation of circuit breakers) need be considered. The current of interest during this time is that which flows out over the power line wires and returns through the ground, called "residual" current. The voltage induced is along the telephone wires in parallel. Since the telephone circuits are metallic, the talking paths over them are usually not seriously affected by the fundamental frequency voltage unless this voltage causes the telephone protectors to operate. Service over grounded telegraph circuits may, however, be impaired even if the induced voltage does not reach values high enough to operate protectors, and the telephone circuits under this condition may be made noisier than usual.

It can be seen that the electrical phenomena in which we are interested will be affected by three basic factors. The first of these is concerned with the "magnetic coupling" between the power and telephone lines, considered with ground return for the reasons pointed out above. This coupling is a function of the strength and frequency of alternation of the flux set up at the location of the telephone line by a given amount of ground return current in the power line. The second factor is concerned with the amount of ground return current in the power circuit at the time of a ground fault since, for a given coupling, this will determine the strength of the magnetic field. The third factor is concerned with the conditions in the telephone plant which determine its reactions to a given induced voltage. Each of these three factors is taken up individually in the following discussion.

Since low-frequency induction between power and telephone circuits involves a series of separate and distinct occurrences, it is evident that

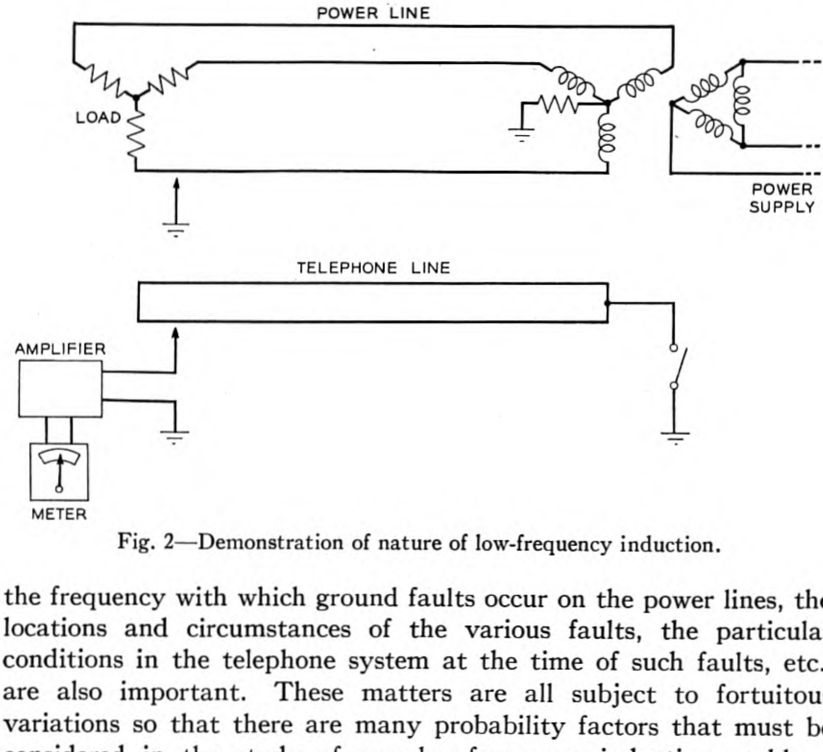


Fig. 2—Demonstration of nature of low-frequency induction.

the frequency with which ground faults occur on the power lines, the locations and circumstances of the various faults, the particular conditions in the telephone system at the time of such faults, etc., are also important. These matters are all subject to fortuitous variations so that there are many probability factors that must be considered in the study of any low-frequency induction problem. These probability factors cannot be demonstrated with the apparatus available.

In order to illustrate the general nature of the electrical phenomena, Fig. 2, which shows parallel power and telephone systems with the power line supplying a load, has been simulated using the miniature lines. The load current may be substantial in magnitude but is normally confined to the line wires and is commonly called "balanced current." Under this condition, the induced voltage is small. If, now, one of the power conductors is grounded, current will flow to ground at that point. This current is the "residual current" mentioned previously. When this is done in the demonstration in such a way that the residual current flows through the exposure, the voltage induced along the telephone circuit rises very materially. Also, since the induced voltage acts along the telephone circuit, opening the ground connection at the far end of the telephone line reduces the voltage shown by the meter to a very small amount.

That the power current causing the induction is unaffected by transpositions in the power circuit, can be shown by transposing the power circuit in the set-up. No appreciable change in the induced voltage occurs when this is done. Likewise the induced voltage, since it is induced along the telephone wires in parallel, is unaffected by telephone circuit transposition, as can be shown by transposing the telephone circuit. Consequently, the matter of power or telephone circuit transpositions can be neglected in the further analysis.

Coupling Factors

Using the demonstration arrangements, some of the basic factors in coupling can be observed. For example, since the voltage is due to magnetic induction and accumulates along the telephone circuit, the coupling should be proportional to the length of the (uniform) exposure through which the fault current flows. This can be observed by noting the reduction in induced voltage as the fault on the power line is moved from the end of the exposure toward the supply end. (In the demonstration, the fault current is the same regardless of the location of the fault.)

Likewise, if the voltage accumulates along the telephone circuit, the longitudinal voltage measured should be proportional to the length of telephone circuit exposed. This can be observed by again placing the fault on the power line at the end of the exposure and moving the measuring point along the telephone line. As this point is moved toward the grounded end, the indicated voltage goes down.

Another basic factor in coupling is its relationship to the separation between the lines. Generally speaking, the greater the separation, the smaller the coupling. How much coupling will exist for a given separation depends on a number of factors, one of which is the structure of the earth. This effect will be discussed first.

In the type of problem we are considering the telephone wires comprise one side of a long loop, the other side of which is the earth. Likewise, the power wires comprise one side of a loop, the other side being the earth. It is a well known fact that the magnetic coupling between two parallel loops at a given separation increases as the sizes of the loops increase. The sizes of the loops in the case being considered are determined by the distribution of the return current in the earth.

A great deal of theoretical and experimental work has been done in connection with the analysis of the distribution of current in the earth, and it has been found that one of the important factors is the "resistivity" of the earth. The effect of resistivity of the earth can be briefly summarized as follows:

(*a*) Considering the outgoing and return paths for residual current on a power line, the mutual induction between the current in the wires and the return current in the earth tends to pull the earth currents together and to concentrate them under the line as near the surface as practicable. This action tends to decrease the coupling to an adjacent circuit by decreasing the effective separation of the sides of the loop.

(*b*) The resistance which the current encounters in flowing through the earth tends to make it spread out because, by so doing, the current density is reduced and the voltage drop is consequently reduced. This spreading out tends to increase the coupling to an adjacent circuit.

(*c*) The net distribution of the current in the earth is a balance between these two opposing tendencies and this distribution will be different for different resistivities of the earth. Generally speaking, the greater the resistivity of the earth, the more the current will spread and the greater will be the coupling to an adjacent circuit.

Figure 3 is a graphical representation of how the return current in the earth tends to spread with an increase in earth resistivity. While this figure shows only the vertical spread, a similar spreading also takes place horizontally.

The effect of the sizes of the primary and secondary loops on the coupling is greater when the loops are widely separated than when they are close together. For this reason, the effect of earth resistivity on coupling is much greater for wide separation exposures than for exposures where the lines are close together. Consequently, with high resistivity earth, the coupling not only is higher at all separations

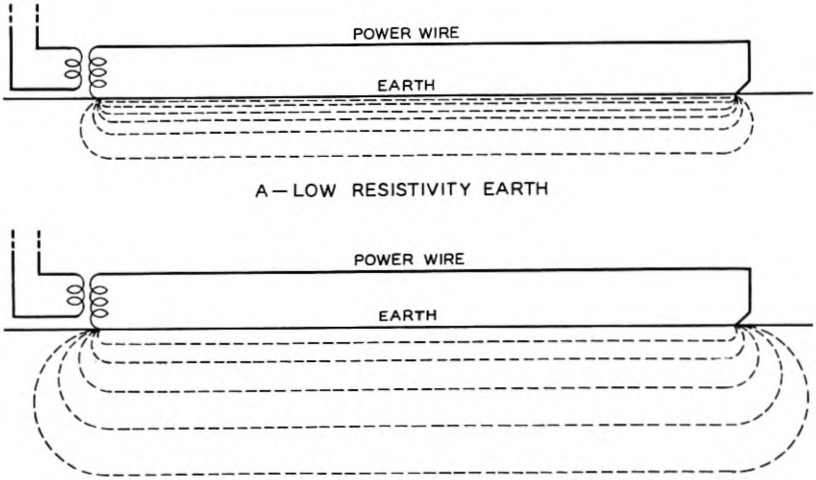


Fig. 3—Current distribution in earth of different resistivities.

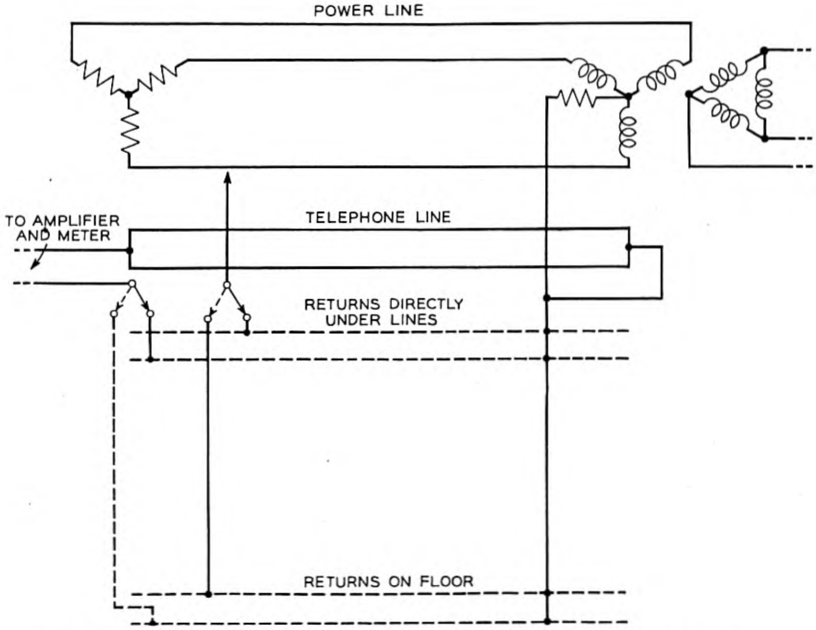


Fig. 4—Demonstration of effects of depth of return current in earth.

than with low-resistivity earth, but (except for very wide separations) the percentage reduction secured by increasing the separation a given amount is smaller.

Figure 4 shows a schematic of a demonstration set-up to show some of these effects. It is, of course, impracticable to employ earth of different resistivities in a demonstration of this kind, but the difference between return current which concentrates near the line and that which is more remote can be roughly indicated. In order to illustrate the fact that when the currents are concentrated closely under the line, the coupling falls off rather rapidly as the separation is increased, returns immediately under the lines are used and the telephone line is moved to change the separation. To illustrate that a wider distribution of current in the earth tends to increase the coupling and to make it less affected by separation, returns on the floor are used for both the power and telephone lines. It can be shown that:

(*a*) The induced voltage increases when the connections are changed from the upper to the lower returns.

(*b*) When the upper returns are used, *the percentage reduction* in induced voltage when the separation is increased, is greater than the percentage reduction when the lower returns are used and the separation is increased by the same amount, i.e., when the telephone line is moved between the same positions of minimum and maximum separation.

If the earth is not homogeneous, that is, has strata of different resistivities, the distribution of the earth current is distorted and varying effects are noted. Where local irregularities exist, marked and sometimes erratic changes in coupling may occur within comparatively short distances. An "effective" earth resistivity can usually be determined by test even where the earth is stratified.

Another important factor in determining the net coupling between power and telephone circuits is the effect of grounded wires or other linear grounded metallic structures along the inductive exposure. Voltages are induced in such grounded metallic structures in the same way as voltages are induced in telephone wires and these voltages cause currents. The magnetic fields accompanying these currents generally oppose those from the power wires and reduce the induction in the telephone circuit. The effect of such currents in grounded structures is generally spoken of as "shielding."

The amount and phase of the current in a grounded conductor in a given location and hence the shielding provided by it depend on the impedance of the conductor with earth return. Hence the shielding is increased when the resistance of the conductor and its ground connections is reduced.

In order to illustrate these effects, the demonstration shown in Fig. 5 has been set up. With the shield wire on the power line, the shielding effect can be shown under two conditions as follows:

(*a*) When the switch directly grounding the shield wire is closed the induced voltage in the telephone circuit goes down materially due to the shielding effect of the current in the shield wire.

(*b*) When, instead of grounding the shield wire directly, it is grounded through a small resistance, the reduction in induced voltage is much smaller because the resistance limits the current in the shield wire.

In order that a conductor may exert a shielding effect, it must have a substantial coupling to either the power or telephone line; i.e., it must be fairly close to one or the other. By moving the wire shown in Fig. 5 it can be demonstrated that:

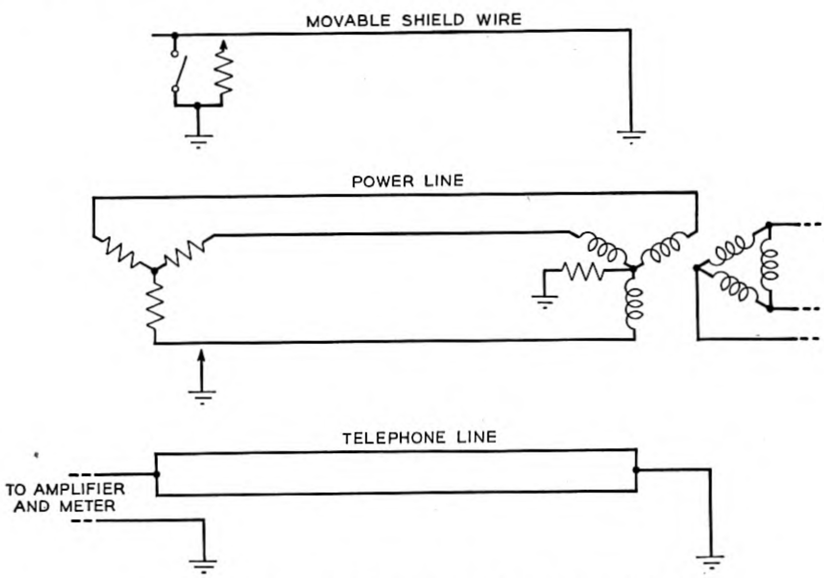


Fig. 5—Demonstration of shielding effect of grounded wire.

(*a*) With the shield wire on the power line or on the telephone line, substantial shielding is secured.

(*b*) If the wire is moved outside the exposure, the shielding is reduced to a small value.

While the demonstration shows only the effect of a grounded wire on or near the power or telephone lines, similar effects in varying degrees may be caused by such grounded metallic structures as underground pipe lines, railroads where the rails are bonded in long lengths, trolley lines, etc. In many situations the shielding effects of such structures may be substantial.

Another type of grounded conductor which may give substantial shielding is the metallic sheath of a telephone or power cable. A cable sheath will effect some shielding on conductors which are not enclosed by it in the same way as any other grounded metallic conductor, but the major shielding effect is experienced on conductors within the sheath. The shielding effect of a cable is, as in the case of a shield wire just demonstrated, determined to a considerable extent by its impedance with ground return.

Shielding due to a telephone cable can be demonstrated using the set-up shown in Fig. 6 and it is noted that:

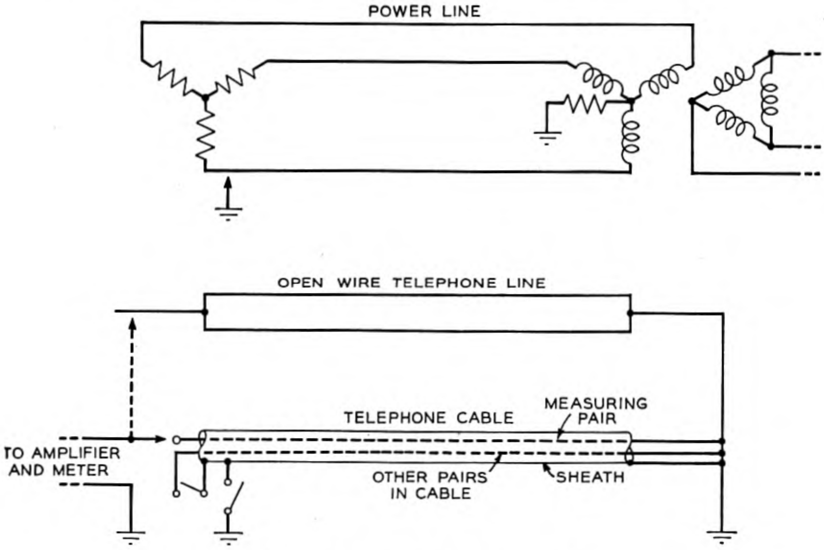


Fig. 6—Effect of cable shielding.

(*a*) The voltage along a conductor inside the cable is reduced when the sheath is directly grounded at both ends.

(*b*) If the effective resistance of the sheath is reduced by paralleling with it some of the conductors inside, the shielding is increased (i.e., the reduction in voltage is greater).

(*c*) When resistance is added in one of the sheath-to-ground connections, the shielding effect of the sheath and conductors is reduced.
(*d*) If the shield wire is grounded its effect is cumulative [2] with that of the cable sheath and conductors.
(*e*) The voltage along the open wires on the line is reduced when the cable is grounded at both ends. Here again, of course, the effect of the shield wire is cumulative [2] with that of the cable sheath and conductors.

The same shielding effects could be shown if the power instead of the telephone circuit were in cable. Also, of course, if there is more than one power or telephone cable, the shielding is increased. Iron armoring also tends to increase the shielding.

Power Circuit Conditions

Having illustrated some of the factors affecting coupling let us now review briefly the factors affecting the current in the power line at

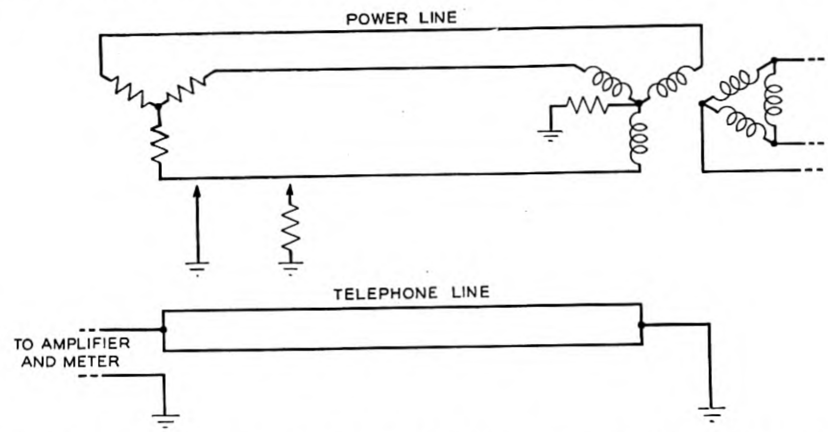


Fig. 7—Comparison of induced voltage with different amounts of residual current.

times of ground fault. Not only is the current flowing to ground under fault conditions on a power line usually greater than the current under normal operating conditions, but it is also a fact that a given current in a ground return circuit induces a larger voltage along a paralleling telephone circuit than the same amount of current if confined to the phase conductors. That the induced voltage depends on the amount of residual current can be demonstrated as shown in Fig. 7. Here a

[2] While the benefits are "cumulative," the individual effects are not directly additive due to mutual reactions between the different shielding conductors, the net effect being less than the total of the effects of each shielding conductor acting alone.

comparison is made between the induced voltage for two residual currents, one greater than the other. With larger current, the voltage indicated on the meter is larger. The resistance inserted in the fault to reduce the current in this demonstration might be thought of as simulating added impedance anywhere in the ground return circuit through the fault. For example, it might be thought of as simulating the effect of the line impedance which would be added if the fault occurred at some distance beyond the end of the exposure. Also, its effect is the same as would be produced by an increase in the reactance of the supply transformers or in the neutral-to-ground connection; or by an increase in the local resistance at the fault itself.

In analyzing the impedances further, there are two general types of power systems which must be considered: the "grounded neutral" and "isolated neutral" systems. These are illustrated in Fig. 8. In the grounded neutral system, the neutrals of one or more transformer banks are grounded directly or through impedance so that in the event of a fault, a path for current is established from the fault through the earth and back to the system through the neutral-to-ground connections. In the isolated neutral system there are normally no grounds on the system so that in the event of a fault the only path for fault currents is through the capacitances of the unfaulted phases to earth or through a second fault if one exists. Hence for a single fault, the fault current is limited to the charging and leakage current.

Figure 8-*D* shows for a grounded neutral system, the equivalent single-phase circuit for residual currents. In an actual line, the circuit conditions are, of course, usually much more complex than those shown. In even the simplest situations, there are usually other lines, generator points, or grounding points which supply some fault current. However, for the purpose of examining the fundamental phenomena, the simplified diagram can be used. As can be seen, the impedances which control the residual current are those associated with the fault, the line impedance, those in the transformer and generating equipment and the impedance, if any, in the neutral-to-ground connection. Impedances in any of these places tend to limit the fault current.

Figure 8-*C* shows a simplified diagram of the equivalent single-phase circuit of an isolated neutral system with a single-phase fault-to-ground. For this condition, it is evident that the fault-current path includes the capacitances to ground of the unfaulted phase conductors. In a small system these capacitances will be small and the fault current will, therefore, also be small, particularly if the voltage is not high. In extensive systems or systems having much

cable, the capacitances and hence the fault current may be fairly large, particularly if the voltage is high. Of course, in an actual system the capacitances to ground are distributed throughout the system so that the amount of residual current in the lines will vary from location to location, being a maximum at the fault and tapering to zero at the end of each branch of the system.

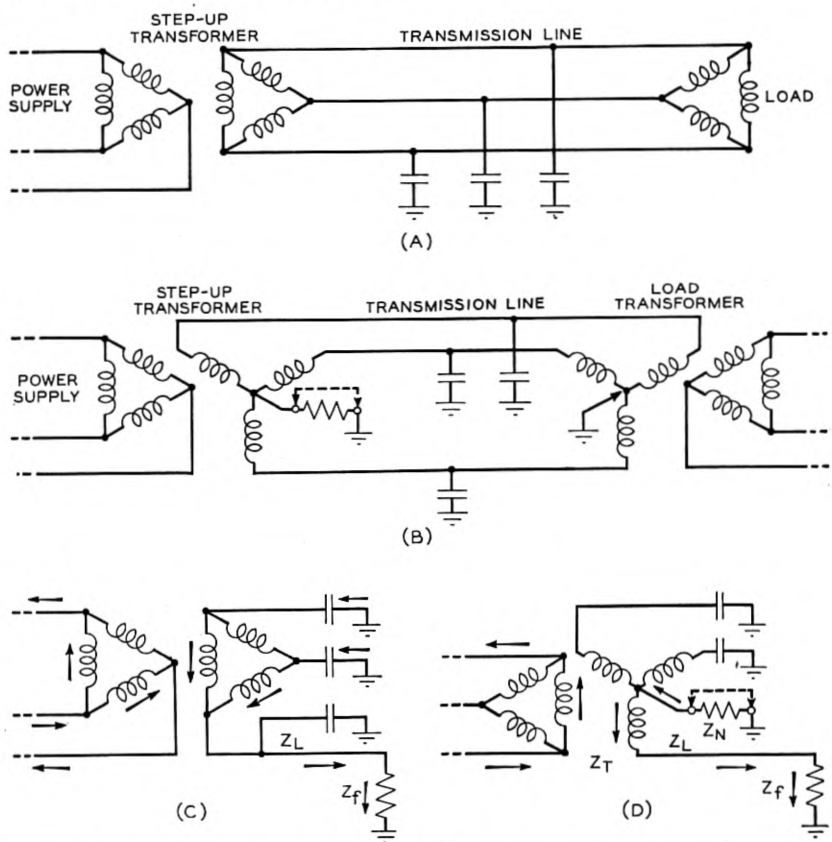


Fig. 8—Types of power systems. (*A*) Isolated neutral system under normal conditions. (*B*) Grounded neutral system under normal conditions. (*C*) Isolated neutral system with single fault. (*D*) Grounded neutral system under fault conditions.

If, in a power system, a second fault-to-ground occurs on another phase while the first persists, a large residual fault current will exist in the line between the faults even if the neutrals are isolated. Simultaneous faults on two phases at different points may occur on any type of system, but are more likely to occur on an isolated neutral system than on one in which the neutral is solidly grounded. This is due to the fact that for the isolated system, full phase-to-phase or possibly higher voltage is impressed between the unfaulted phases and ground, thus increasing the voltage stress on the insulation of the entire system during the time of fault.

Figure 9 shows a demonstration set-up to illustrate the effects of faults on an isolated neutral system which is small enough so that the capacitances are negligible. It will be noted that when a single fault

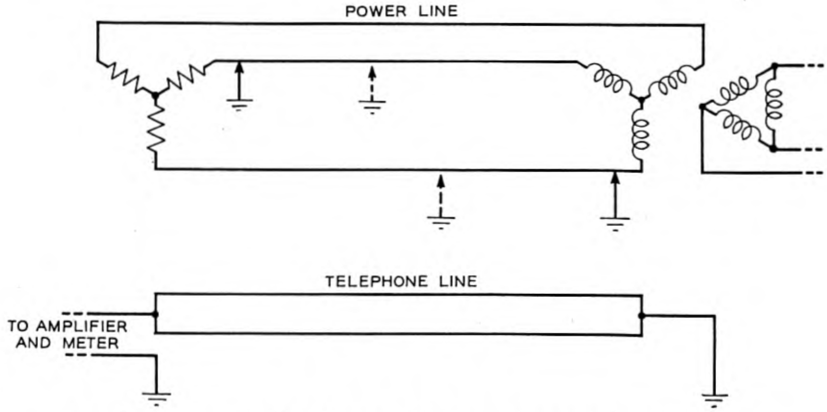


Fig. 9—Demonstration of effect of faults on isolated neutral system.

is put on the line at the far end, no appreciable rise in voltage on the telephone line occurs. However, when a second fault is put on another phase at the near end, the induced voltage immediately rises. In order to illustrate that the current is residual only between the faults, the faults can be moved closer together. When this is done, the induced voltage decreases until, with the faults at the same point, it practically disappears. The small remaining voltage is largely due to balanced current induction due to the heavy load on the power line when two of the phases are shorted.

A system grounded through a neutral impedance, such as when resistance or reactance is included in the neutral-to-ground connections or when a high reactance grounding bank is used, partakes of some of the characteristics of an isolated neutral system. Generally speaking, the addition of neutral impedance tends to reduce the fault current, this effect being proportionately larger for faults near the neutral grounding points. This reduction in fault current tends to reduce the voltage induced on nearby telephone lines and in some cases may reduce the "shock" to the power system and the damage at the point of fault. On the other hand, increasing the neutral impedance tends to increase the difficulty of securing adequate selectivity in power system relay operation and to make a more complex relay system necessary. It also tends to increase over-voltages on the power system and to reduce the factors of safety for lightning arrestors.

In addition to the magnitude of the residual current, its duration is of importance since the length of time that the induced voltage persists on a telephone circuit has important reactions on its effects. For example, the chance of permanently grounding telephone protectors, with consequent interruption of service until the protector blocks are replaced, depends not only on the amount of current through the blocks but upon its duration. Likewise, many of the other effects, which are described later, are materially affected by the duration of the induced voltage. Since, except for self-clearing faults, the duration of fault current is determined by the time of operation of the relaying system, the reliability and speed of operation of the latter is an important factor. There are many types of relaying systems and it is not practicable to go into a discussion of them here except to point out that rapid and reliable relaying is usually simplest on a solidly grounded neutral system. For systems with large impedances in the neutrals, it may be difficult to secure rapid fault clearance, particularly if the system layout is complicated. For isolated neutral systems, rapid relaying on ground faults may be very difficult or impracticable.

Telephone Circuit Conditions

The voltage due to magnetic induction accumulates *along* the telephone circuit and can be represented as a voltage in series with the telephone wires. Figure 10 shows schematically how this voltage acts. The two sides of the metallic telephone circuit are assumed to have the same induced voltage and impedance and are shown here replaced by a single equivalent conductor. The total voltage which is equal to the product of power line fault current and coupling is represented by a number of generators connected in series through impedances representing, in total, the line impedance inside the exposure. At the ends of the exposure are connected impedances representing those in the line and between line and ground outside of the exposure. The longitudinal induced voltage acting through the series and shunt impedances of the telephone line will produce the following conditions of interest:

(*a*) Voltages between the telephone wires and ground at various places along the telephone line.

(*b*) Current in the longitudinal telephone circuit.

(*c*) Voltages between different wires on the telephone line.

Telephone circuits are supplied with protective devices. The part of the telephone protective system of most interest in connection with low-frequency induction is the carbon-block protector. This device provides a small air gap between carbon surfaces one of which is connected to the telephone conductor and the other to ground or cable sheath. When an excessive voltage is impressed on the telephone

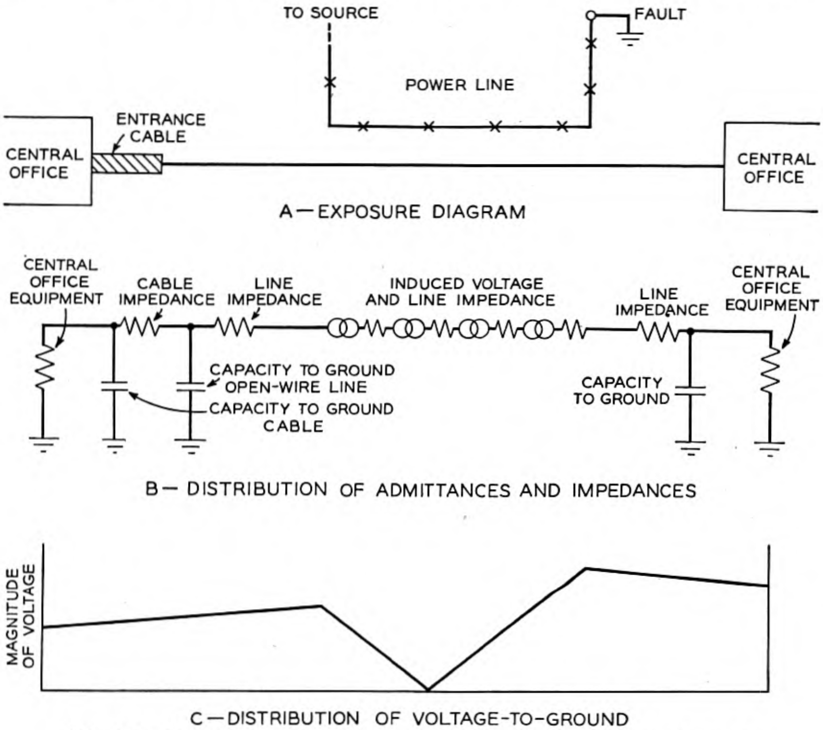


Fig. 10—Distribution of voltage to ground before protectors operate.

conductor an arc is established in the air gap thereby grounding the conductor. Protectors are located at central offices and at other points, such as at junctions of open wire and cable where it is desired to limit voltages on telephone wires due to lightning, contacts with power systems, induction, and other extraneous voltages.

In analyzing the distribution of induced voltage between a telephone circuit and ground assume first that no protectors are operated. Under this condition, the voltages to ground on the telephone wires at various points are determined by the impedances between the wires and ground along the line and at central offices where equipment is connected to them. The voltage to ground at either end of the exposure is equal to the longitudinal current times the impedance-to-ground seen looking away from the exposure at that end. Figure 10-*C* illustrates how the voltages may distribute due to the distribution of impedances between the wires and ground along the line and in the central office equipment. Of course, in practice, the variety of impedance distributions encountered is almost infinite and the corresponding voltage distributions vary over a wide range.

If voltage-to-ground at any point where protectors are located exceeds the operating voltage of the protector, the protector operates and three things happen:

(*a*) The voltage-to-ground at the place where the protector operates is reduced to a low value. This makes the longitudinal voltage pile up at the protectors at the opposite end, and in most cases, they will also operate.

(*b*) The operation of the protectors at the two ends completes a loop consisting of the telephone circuit and ground so that the induced voltage will cause current to flow through both protectors.

(*c*) The voltages-to-ground on the circuits on which protectors have operated are changed and redistributed and the voltages on the other telephone circuits are also changed and redistributed due to shielding, as discussed later.

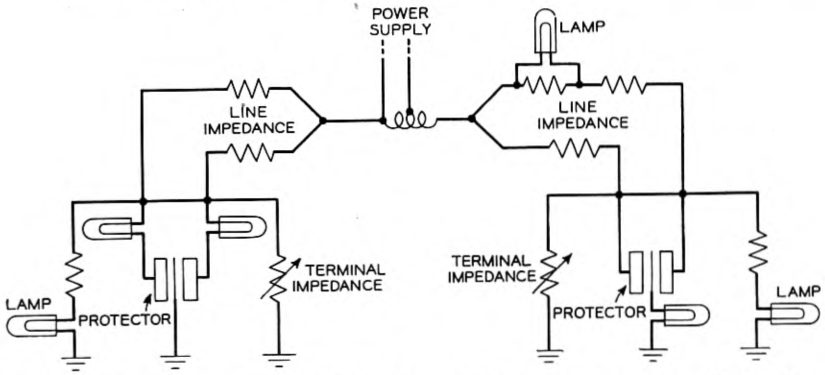


Fig. 11—Demonstration showing effect of terminal impedance on voltage distribution.

All of these effects take place within a very short time after the longitudinal voltage is applied so that for all practical purposes they can usually be considered as being instantaneous.

In order to illustrate these phenomena, the demonstration shown in Fig. 11 can be used. In this demonstration the longitudinal voltage is impressed by means of a power transformer in order to secure a sufficiently high voltage. A general idea of the relative magnitudes of the voltages at the two ends can be had by observing the brilliancy of the voltage measuring lamps. By varying the slide wires which control the terminal impedances the proportions of the total voltage which appear at either end are changed and an idea of the changed distribution can be obtained by observing the changing glow of the measuring lamps. Finally the voltage at one end can be increased enough to cause the protectors at that end to operate, whereupon the measuring lamp goes out and the small protector lamp lights. Immediately the other protectors operate, as evidenced by the voltage lamp going out and the protector lamp lighting, and the line current increases as evidenced by the brilliance of the line current indicating lamp.

An important factor in the further analysis is the characteristics of the telephone protector. The arc takes place between two carbon surfaces. The gap between these two surfaces has a very high breakdown speed and a very low impedance after it is broken down. Another important characteristic from the standpoint of low-frequency induction is its tendency to become permanently grounded if heavy currents are discharged or if the discharge continues for some time. Consequently, the amount of current in the longitudinal circuit in the event of a breakdown and its duration are important factors in determining the chance of permanently grounding the protectors and causing the circuit to become inoperative until the blocks are changed. Duration is, of course, ordinarily a function of the duration of fault current on the power line as pointed out previously.

The amount of current through operated protectors is determined by the longitudinal voltage and the longitudinal impedance of the telephone circuit. If, for the moment, it is considered that only one wire is present, this current is simply the total longitudinal voltage divided by the total series impedance of the wire plus any resistance in the protector grounds. This can be seen from Fig. 11.

Ordinarily there are numerous circuits on an open-wire telephone lead or in a telephone cable. If the protectors on a number of these circuits break down, the current in each wire will be less than that which would exist were only one wire present as in the above illustration. This is due to the mutual impedance between the different telephone wires which causes the current in any one wire to reduce the current in the remaining wires. The *total* current in all of the wires of course increases as the number of wires on which protectors have operated is increased but not in direct proportion. Figure 12 illustrates this effect. The actual circumstances concerned in this phenomenon are, of course, that the wires which become grounded at their terminals through the operation of the protectors exert a shielding effect in exactly the same way as any other grounded conductor.

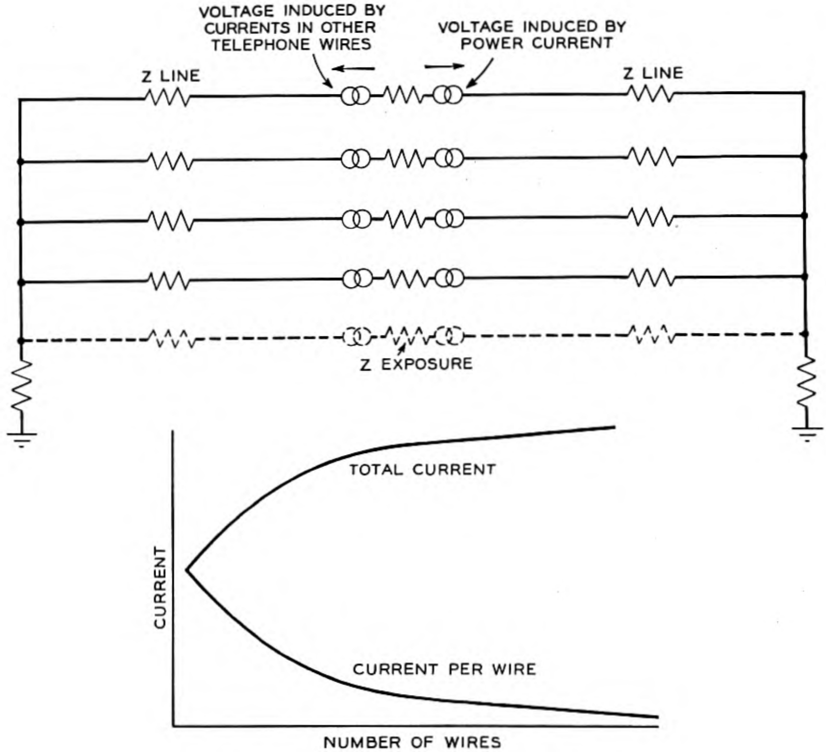


Fig. 12—Current in telephone wires.

The resistance of an individual wire is relatively high but if numerous protectors operate, the shielding may become fairly great, due to the closeness of the wires to each other and to the fact that a substantial amount of copper may be involved. Of course, this shielding is obtained at the expense of at least momentary interruption of the circuits on which protectors operate.

The shielding effect of current in grounded telephone wires is exerted on all telephone wires on the line regardless of whether the protectors on them have or have not operated. Consequently, what may happen on a large telephone line with a moderate induced voltage on it is that enough telephone protectors on different circuits operate to give a shielding effect on the remaining wires sufficient to reduce the voltages on them to values lower than will operate the protectors.

Another important factor is the voltage-to-ground at various places along the telephone circuit after protectors operate. With the protectors operated the voltage-distribution-to-ground can be evaluated from the longitudinal induced voltages, the longitudinal currents, and the series impedances in the circuit. As the simplest and perhaps most striking case, consider a telephone circuit which is solidly grounded due to operated protectors at the ends of a uniform exposure with a fault on the power line at the end of or beyond the exposure. This situation is illustrated in Fig. 13. The distribution of induced

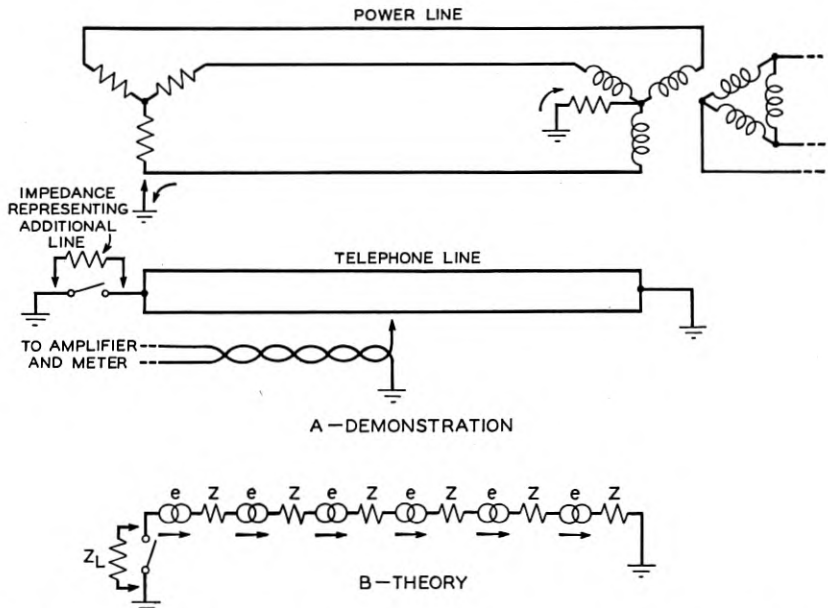


Fig. 13—Voltage to ground for telephone line grounded at ends where fault is outside of exposure.

voltage along the line is uniform and the longitudinal current is equal to the total longitudinal voltage divided by the total series impedance. If the net drop in voltage is taken from either end to any point along the circuit, it will be found that the induced voltage accumulated over this distance is equal and opposite to the voltage drop over this same distance due to the current flow through the impedances in this section. Consequently, under these conditions the voltage-to-ground is zero at all points along the circuit. This is true regardless of the magnitude of the induced voltage. Figure 13 also shows the set-up by which this fact can be demonstrated. It will be noted that, while there is a fairly high longitudinal voltage, the voltage between the wires and ground with both ends of the circuit grounded is negligible at all points along the line.

If the telephone line extends beyond the exposure, the effect of this portion of the line is to add impedance between the exposure terminal and the protector without adding a corresponding induced voltage. If a power line fault occurs at the end of the exposure, a voltage-to-ground will exist at this point equal to the current in the telephone

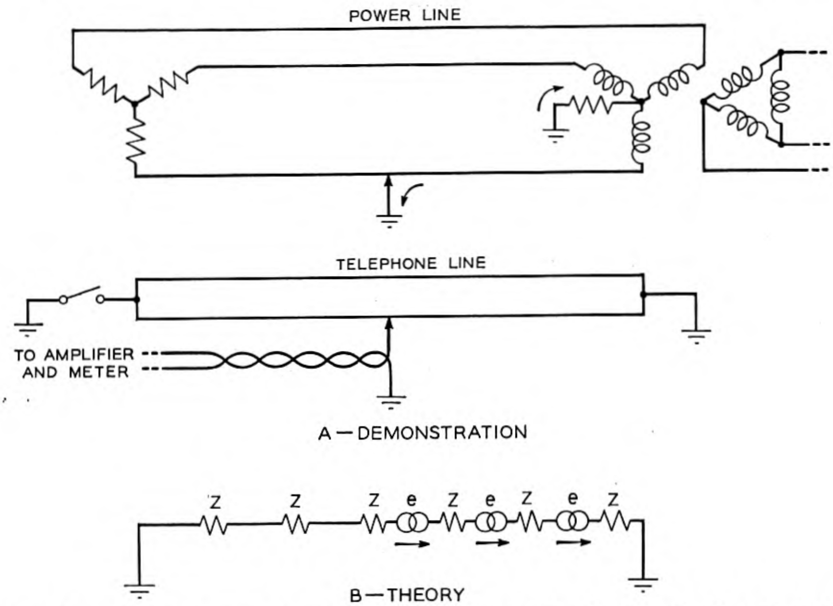


Fig. 14—Voltage to ground for telephone line grounded at both ends where fault is inside of exposure.

line times the impedance outside the exposure. This situation can be illustrated in the set-up of Fig. 13 by leaving the power fault at the end of the exposure and adding a small impedance in the ground connection to the telephone line at one end. When this is done it will be noted that a voltage-to-ground exists at the end where the impedance is added and that the voltage-to-ground decreases as the measuring point is moved toward the other end.

To illustrate that voltage-to-ground may occur under other conditions even though the telephone line is solidly grounded at the exposure terminals, consider the situation shown in Fig. 14. This represents the same exposure conditions as in the preceding set-up and the telephone circuit is solidly grounded at both ends, but the fault current instead of flowing through the entire exposure flows through only half of it. In this case the telephone circuit impedances are the same as in the preceding case, but only half of the induced voltage is present. Consequently, the amount of current through the longitudinal circuit is only half of that in the preceding case. Now if the net voltage drop from either end to the middle is taken, it will be found that it is equal to one-half of the total longitudinal voltage induced under the conditions shown. Figure 14 also shows the set-up for demonstrating this condition. In this case the longitudinal voltage is smaller than in the preceding demonstration, but if both ends are grounded and the voltage measuring device is moved along the line, the voltage-to-ground increases from one end to the middle and then falls off from the middle to the opposite end.

In the last two demonstrations, the fault current on the power line was fed from one end only, i.e., "single-end feed." It sometimes happens that the fault current may be supplied to a power line, at least during the initial stage of a disturbance, from both ends, i.e., "double-end feed." The double-end feed condition tends to reduce the overall longitudinal induction when the fault occurs inside the exposure. In the demonstration shown in Fig. 15 it may be observed that for the set-up with a fault at the middle of the exposure the symmetry is so good that the total longitudinal voltage is very small. However, with the telephone circuit grounded at both ends, a substantial voltage-to-ground exists at a point in the telephone circuit opposite the fault and this voltage-to-ground reduces to zero at the ends. As the fault is moved toward either end of the exposure, there is a tendency for the longitudinal voltage to increase and for the voltage-to-ground, with both ends of the line grounded, to decrease until the limiting condition brought out in Fig. 13 is reached.

The analysis of voltage-to-ground can be carried out for any combination of impedances and induced voltage distributions by totaling vectorially the voltage drops (including any voltage drop over protector ground resistance) and the induced voltages between a grounded point and the point at which the voltage-to-ground is desired. The same analysis can also be carried out regardless of whether one or numerous wires are involved as long as all of the wires are grounded directly or through arrestors at the same points. If some of the wires on a line are not grounded, i.e., the protectors are not operated, the analysis for these wires must be carried out on the basis of their admittance-to-ground as discussed previously. In such a case, the longitudinal voltage to be employed would be that remaining after correction for the shielding effect due to wires on which protectors have operated.

As mentioned previously, mutual shielding of the telephone wires on a line may prevent the operation of some protectors, particularly on a large line. Consequently, the exact analysis of the distribution of voltage-to-ground of all of the wires on a large line becomes very complex. Moreover, as is often the case, if impedance conditions are not uniform, such as where circuits are not coterminous, the complete analysis of the voltage-to-ground becomes even more complicated. Under such conditions it will generally be found that the distribution of voltage-to-ground along the different circuits is different, and consequently, voltages exist between different wires due to these differences in the voltage-to-ground. Likewise, voltages may exist between wires on which the protectors have operated and wires on which the protectors have not operated.

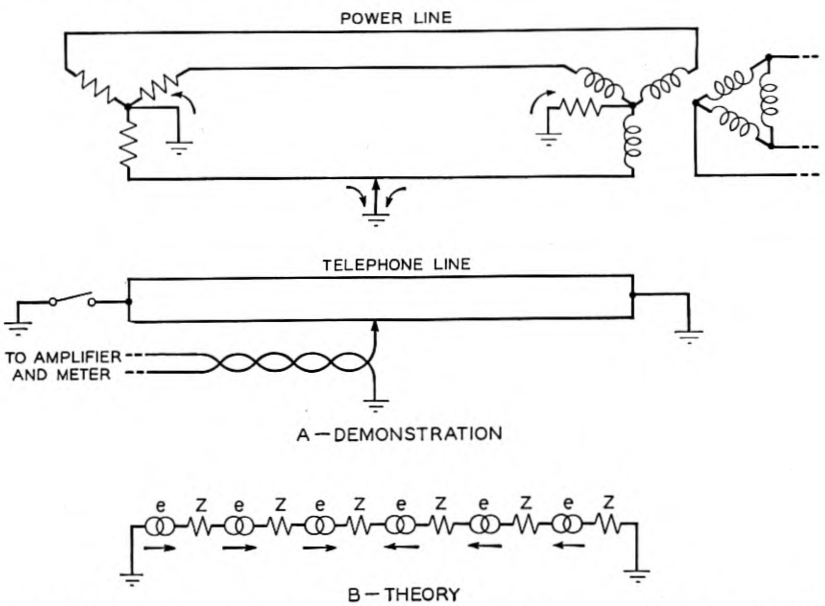


Fig. 15—Voltage to ground for telephone line grounded at both ends where fault is inside exposure—double-end feed.

All of the above analyses have been made on the assumptions of continuous telephone wires. If a wire is opened at one point the longitudinal voltage, reduced by shielding from any currents which exist in other continuous wires, will appear across the "open." On a large line this shielding may, as pointed out previously, be so large that the voltage across the open is reduced to a fraction of the induced voltage.

Using the high voltage equipment which was used in connection with the demonstration of protector operation, "acoustic shock" can be demonstrated. Although strictly the term "acoustic shock" should be used only with reference to the effect on a person subjected to an abnormally loud sound, the term has also come to be used to designate a noise (usually transient) in a telephone receiver, the intensity of which is considerably higher than that of speech. It is produced by an excessive voltage across the terminals of the receiver.

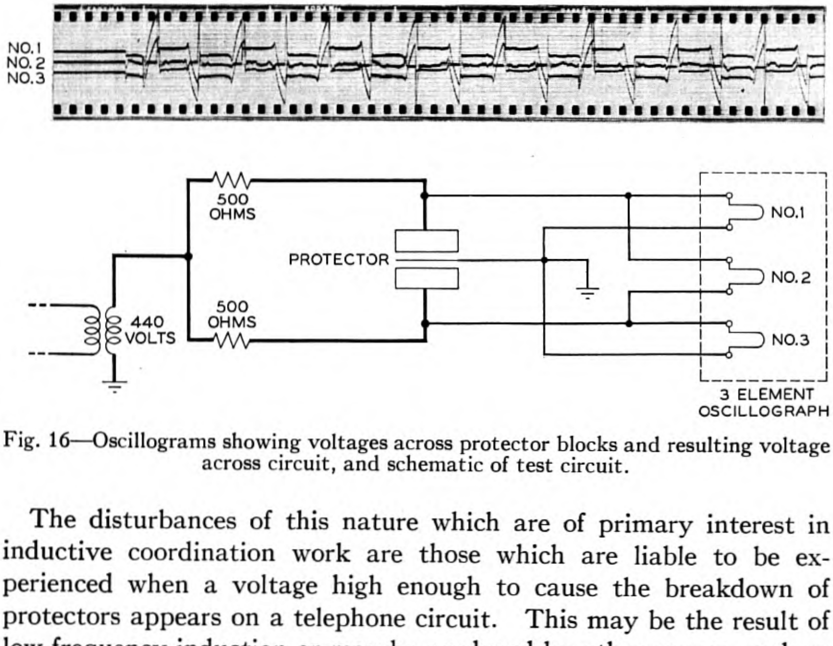


Fig. 16—Oscillograms showing voltages across protector blocks and resulting voltage across circuit, and schematic of test circuit.

The disturbances of this nature which are of primary interest in inductive coordination work are those which are liable to be experienced when a voltage high enough to cause the breakdown of protectors appears on a telephone circuit. This may be the result of low-frequency induction or may be produced by other causes, such as lightning or contacts between power and telephone circuits. Although induced voltages usually appear in equal magnitudes on the two sides of a circuit, the protector gaps on the two sides of the circuit discharge in an unsymmetrical manner with the result that a voltage higher than normal appears across the circuit. When this occurs a loud noise or rattle is produced in the receiver of a telephone set bridged across the circuit.

Figure 16 shows oscillograph traces of voltages measured across operating protector blocks. Each outside trace shows the voltage across one of the two blocks. It will be noted that the two traces are not identical. The middle trace shows the resulting voltage across the circuit. It is this voltage which may cause acoustic shocks. The very jagged outline of this trace indicates that many frequencies other than 60 cycles are present.

The demonstration of Fig. 17 can be arranged to produce acoustic

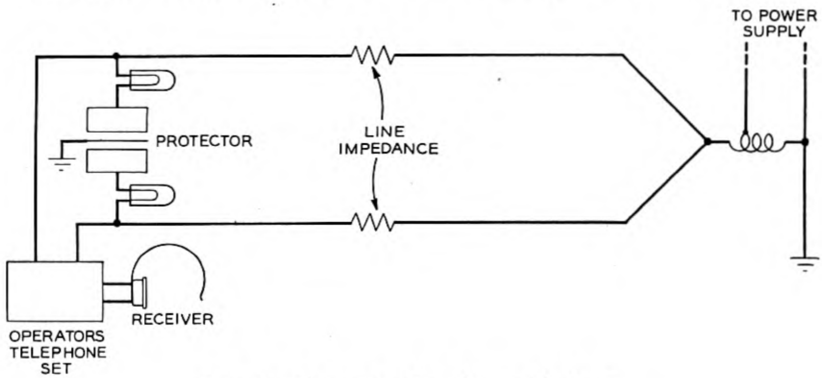


Fig. 17—Demonstration of acoustic shock.

shock. In this demonstration, sufficient voltage is impressed on the circuit to operate the protector blocks at one end. An operator's telephone set and a receiver are connected across the circuit at this end. When the voltage rises high enough to operate the blocks a relatively loud sound is emitted by the receiver.

Other effects may accompany the unsymmetrical discharge of the protector blocks. For example, the signals which are connected at the ends of the telephone circuits may operate and give what is commonly called a "false signal."

Probability Factors

In the preceding discussion, a number of factors were mentioned which may vary between different occurrences in the same inductive exposures. Among these may be mentioned the following:

(*a*) The impedances in the faulted circuit may vary between occurrences due to variations in the location of faults, variations in the effective fault resistance, etc. The effect of the variation in location of the fault, of course, is to change the line impedance in the faulted circuit and hence the fault current.

(*b*) The duration of the fault current may vary between occurrences due to variations in conditions which affect the speed of operation of the power circuit relays and circuit breakers. In some cases, of course, faults may clear themselves without circuit breaker operation and this introduces additional variations.

(*c*) Large variations in longitudinal voltage, voltage-to-ground, and current through protectors may occur with relatively small variations in the locations of faults when they occur inside inductive exposures.

(*d*) The shielding effects due to the operation of protectors on telephone circuits may vary considerably between different occurrences.

The variations in induced voltage duration, etc., between different occurrences are, of course, only part of the story. Obviously, the *total* number of faults which may occur on a power line in an exposure over a given period is equally important. This will be affected by numerous factors such as type of line, severity of lightning and other hazards, etc.

In addition, there are variations in the reactions on the telephone circuits. For example, the protector blocks used do not all break down at the same voltage and the fortuitous variations in the breakdown voltage may have an important bearing on the number of protectors which operate and consequently on the total shielding, current through protectors, etc. From the standpoint of possibilities of acoustic and electric shock, there are of course many other probability factors involved.

All of these factors are under investigation and our knowledge of them is increasing from day to day. It is probable, however, that low-frequency induction will always remain a subject in which quantitative analyses can tell only a part of the story.

Circulating Currents and Singing on Two-Wire Cable Circuits

By LEONARD GLADSTONE ABRAHAM

One of the important factors limiting the working net losses of two-wire cable circuits is the possibility of excessive circulating currents or actual singing.

A theory is developed for the computation of the distribution of singing margins on groups of two-wire circuits from the known gains and losses and known functions of the deviations of loading coils, loading coil spacing, cable capacitance and office equipment. The distribution functions of circulating current margins, of active return losses and of active singing points are also derived.

The possible application of these methods to specific problems is discussed and an example of the computations involved is given.

The theory herein involves certain approximations and empiricisms in determining the singing limitations but it is believed to give an answer which approaches the exact answer rather closely.

Introduction

AMONG the considerations which limit the minimum working net losses [1] of two-wire cable circuits, one of the most important is the desirability of avoiding excessive circulating currents. These circulating currents may manifest themselves as a quality impairment due directly to frequency and phase distortion or as sustained oscillation (singing).

In a given two-wire cable circuit, if the exact location and nature of each irregularity were known, it would be possible to compute exactly whether sufficient singing margin is available. The practicable method, however, is to compute the singing margins which will be exceeded on various percentages of a large group of such circuits, from the information which is available about the irregularities on a distributional basis. This paper first derives theoretical distributions of circulating current margins and singing margins without regard to various practical considerations such as the effect of repeating coils and other apparatus. In the second main division are discussed various considerations which are involved in applying the theory. In the third main division detailed computation methods are illustrated. The attached appendices cover the mathematical derivation of certain quantities.

[1] "Certain Factors Limiting the Volume Efficiency of Repeatered Telephone Circuits," L. G. Abraham, *Bell Sys. Tech. Jour.*, October, 1933.

Theoretical Distributions

Circulating Current Margins

A two-wire loaded cable circuit will be considered which consists of a number of repeater sections with 22-type repeaters. As shown on Fig. 1 there are various circulating paths in such a two-wire circuit

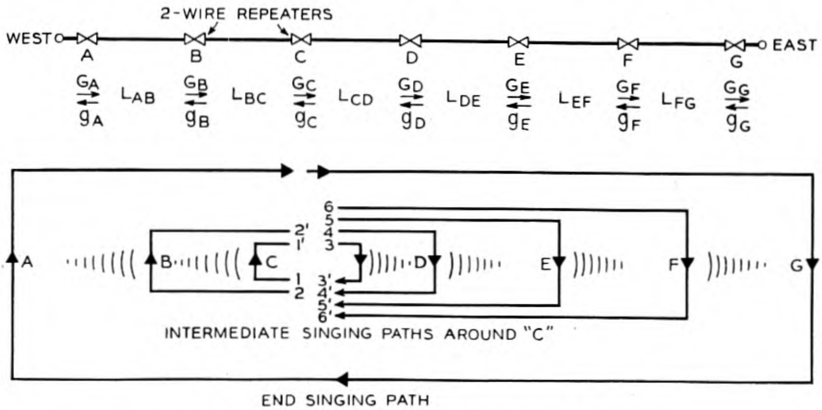


Fig. 1—Singing paths in a two-wire circuit.

which might cause objectionable circulating currents at any given frequency. Each repeater section consists of a large number of loading sections which have approximately the same capacitance and the same inductance per mile. Practically, however, there are deviations of the capacitance and inductance in a given loading section from the nominal average, each of which introduces an impedance irregularity which prevents the balancing network from exactly balancing the line. To determine the amount of current returned from each of these irregularities in a given case and the phase at which it is returned is, in general, impracticable. It is possible, however, to determine in what percentage of circuits the circulating current at any given frequency will exceed a certain percentage of the original current or, in other words, the percentage of cases in which the loss in the circulating current path will be less than a certain amount. The loss in the circulating current path at a given frequency is called the circulating current margin at that frequency.

The distribution of return losses at any given frequency in a cable section without repeater has been determined by G. Crisson in a paper entitled " Irregularities in Loaded Telephone Circuits," *Bell System Technical Journal*, October, 1925. This derives the distribution of the return losses which would be measured at a given frequency on a large number of such cable circuits in terms of certain functions of the capacitance and inductance deviations. From this paper the distribution function of the return losses may be expressed in decibels as follows:

$$S = S_H + S_w - S_A + S_F = S_1 + S_F$$

(See Appendix VI for Nomenclature).

In this equation S_F is the distribution function of the return losses and S_1 is the return loss at the frequency in question which is exceeded by

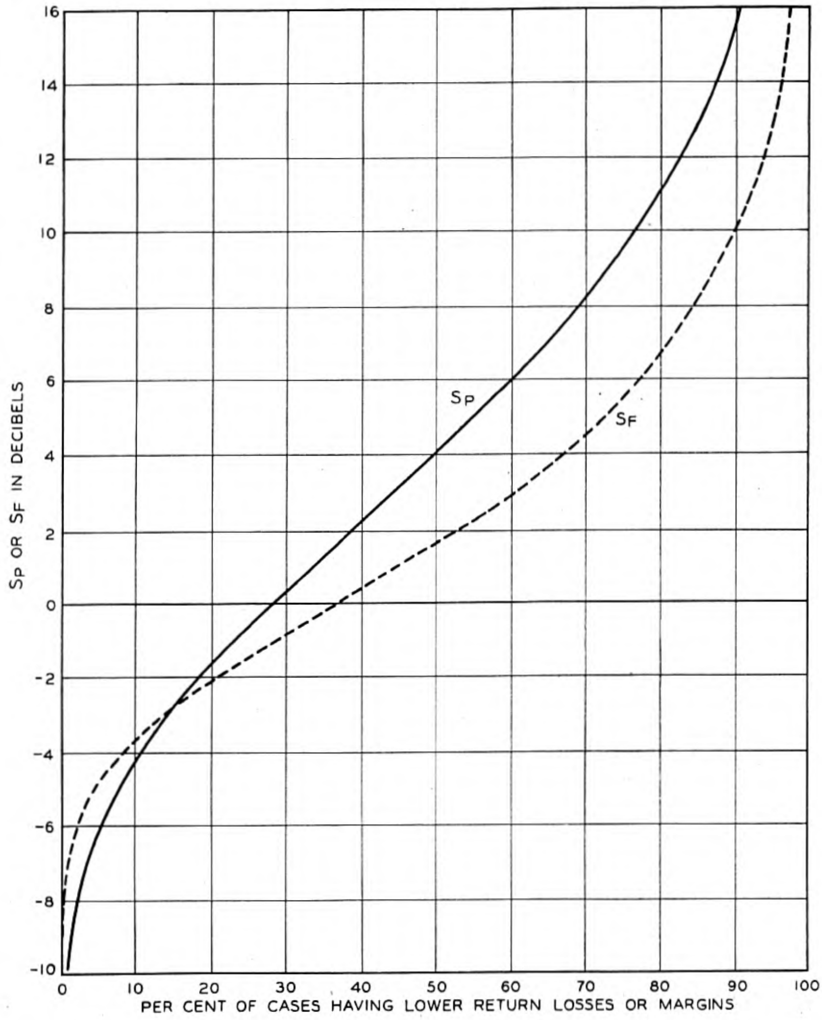


Fig. 2—Values of S_P and S_F.

63 per cent of the total return losses measured. S_H is determined entirely from the fractional deviations of the capacitances and inductances. S_w is determined by the proximity of the frequency in question to the cutoff frequency of the line facilities. S_A is determined by the attenuation of the cable circuit and is, in effect, the summation function of the different sources of irregularity.

Figure 2 shows the distribution curve S_F from the paper referred to above. In a similar manner the distribution curve of several repeater sections in tandem or in parallel [2] may be determined. This is known as an active return loss. Referring to Fig. 1 the return loss measured across the west hybrid coil of repeater C will be determined not only by the irregularities in the immediately adjacent repeater section but also by the irregularities in the other repeater section to the west of repeater C as seen through the intervening losses and gains of the circuit. Strictly speaking, the return losses on the east sides of repeaters A and B will also affect the active return loss because circulating paths around each repeater and around various combinations of repeaters will exist. However, these circulating paths have so much loss in any practical field circuit in which the other requirements are satisfied, that the return losses on the east side of A and B need not be considered.

Appendix I derives the formula for the distribution curve of an active return loss when the passive return losses of which it is made up are of the form given above, assuming no returned currents from beyond the terminal repeater. This distribution function is:

$$S_1 - F(N, T) + S_F.$$

In this case it is assumed that the value of S_1 is the same for the passive return losses of each repeater section, but from the appendix, the more general case where there is a different value of S_1 for each section may be determined.

Figure 3 gives values of $F(N, T)$ for the specific case where each repeater section has the same loss and each intermediate repeater has the same gain. The value of this function is also derived in Appendix I.

Referring again to Fig. 1 there will be an active return loss which may be measured on the west side of repeater C and also an active return loss which may be measured on the east side of repeater C. Assuming that the distribution function of the active return loss toward the west is $S_{11} + S_F$ (not including path A) while the dis-

[2] While this paper develops the theory specifically for the case of repeater sections in tandem, e.g., in a single two-wire circuit, it is generally applicable also to the case of repeater sections in parallel, e.g., in a toll conference connection.

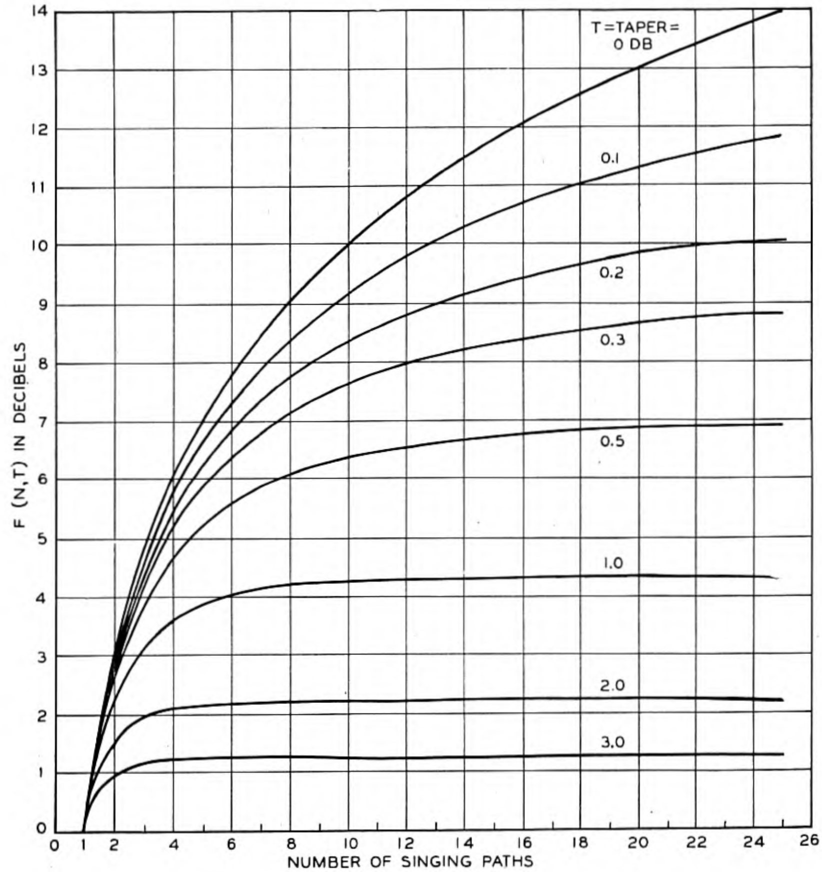


Fig. 3—$F(N, T)$ in perfectly regular circuit.

tribution function of the active return losses toward the east is $S_{12} + S_F$ (not including path G), the distribution function of the two return losses in series around the repeater will be $S_{11} + S_{12} + S_P$ as derived in Appendix II. In the general case, there will be gain in series with the active return losses, which may be considered as part of the return losses. For convenience, define S_{11} as follows:

$$S_{11} = S_1 - F(N_1, T) - G_C. \tag{1}$$

S_1 and $F(N_1, T)$ are as discussed above; N_1 is the number of repeater sections towards the west end of the circuit; and G_C is the gain of the repeater in the west to east direction, say. Similarly,

$$S_{12} = S_1 - F(N_2, T) - g_C. \tag{2}$$

The various quantities are like those enumerated above except that they are for the active return loss in the other direction and include the gain of the repeater in the east to west direction.

Figure 2 shows the value of S_P as derived in Appendix II plotted with S_F. By comparison of these curves, it may be seen that the spread of active return losses around a repeater will be somewhat larger than the spread of active return losses in one direction. The S_P curve approaches more nearly to a so-called normal law than the S_F curve does.

The previous discussion has been confined to the returned currents obtained from intermediate points in the circuit and also at a single pre-selected frequency. In addition to these currents, there is at each end of the circuit a current returned through the path called the terminal return loss. In a four-wire cable circuit this is the only path for circulating currents. From field data it appears that these terminal return losses approximately follow the S_F distribution curve also. The end path toward the west may be written as:

$$S_{21} + S_F = 7 + E_1 + S_F, \tag{3}$$

where 7 db is assumed as the terminal return loss for $S_F = 0$. In this case E_1 is the net loss at the particular frequency from the repeater in question (from the west side of repeater C) to the far end of the circuit and from that end of the circuit back to the repeater minus the gain of the receiving repeater on the latter side of the circuit (the W–E repeater). The end path toward the east is

$$S_{22} + S_F = 7 + E_2 + S_F. \tag{4}$$

If the losses of each cable section are equal and the gains of each intermediate repeater are equal, the repeater in question is at the center of the circuit, and $E_1 = E_2$, and at 1000 cycles E_1 would be equal to the nominal circuit net loss.

The active return loss at a given frequency toward the west which includes the end path (both including the west to east repeater gain) will be

$$S_{31} + S_F = (S_{11} \underset{p}{\mathrm{X}} S_{21}) + S_F \tag{5}$$

and toward the east will be

$$S_{32} + S_F = (S_{12} \underset{p}{\mathrm{X}} S_{22}) + S_F. \tag{6}$$

Where $\underset{p}{\mathrm{X}}$ means that the quantities so connected are combined as if their powers added directly; e.g., $S_{11} \underset{p}{X} S_{21} = S_{31}$ means that

$$10^{-S_{11}/10} + 10^{-S_{21}/10} = 10^{-S_{31}/10}.$$

The circulating current margin (M_c) around this repeater at a given frequency will therefore be $S_{31} + S_{32} + S_p$, as derived in Appendix II. Written in the more general form, the circulating current margin is

$$M_c = [S_1 - F(N_1, T) - G_C] \underset{p}{X} [E_1 + 7] + [S_1 - F(N_2, T) - g_C] \underset{p}{X} [E_2 + 7] + S_P. \quad (7)$$

Singing Margins

The objectionable effects of too low circulating current margin are to cause poor quality in transmission over the circuit and to cause the circuit to " ring " or sound " hollow " to the talker or listener. When a circuit oscillates or " sings," conversations over it become difficult or impossible, voice-operated devices on connecting circuits are locked up, parts common with other circuits such as common " C " batteries may be adversely affected, and other circuits are made noisy through crosstalk in the cable or in the repeater station. It is therefore important that the percentage of cases in which singing occurs shall be very small.

In general, the tendency will be for most of the loaded cable circuits to sing within a fairly limited frequency range which is usually near the upper frequency point at which the overall circuit loss begins to be appreciably greater than the loss at 1000 cycles. Figure 4 shows a

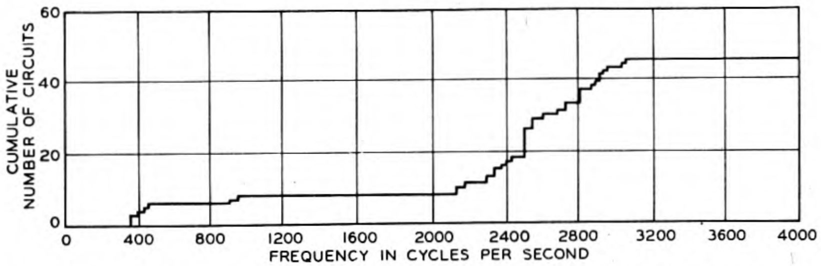


Fig. 4—Singing frequency in 22 tests on 46 19-gauge B & H–88–50 two-wire cable circuits.

cumulative plot of the singing frequencies during 22 tests[3] on 46 19-gauge B & H–88–50 two-wire cable circuits. It may be seen that

[3] A 22 test is a singing test made by increasing the gain of a normal working repeater in a two-wire circuit until singing begins.

over 80 per cent of these frequencies were between 2200 and 3100 cycles, while the nominal transmitted band on these facilities is from 250 to 3000 cycles.

Consider the case of the return loss of one circuit in a group of cable sections. A typical curve of the passive return loss of such a circuit is shown on Fig. 5. It will be seen that it consists of a set of "wabbles"

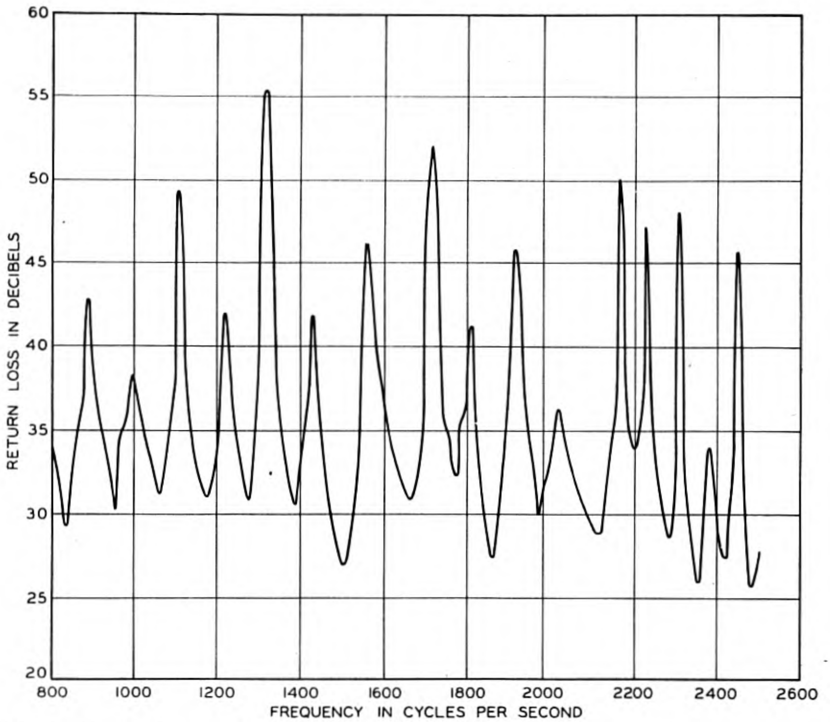


Fig. 5—Typical return-loss-frequency characteristic of a 19-gauge H–88–50 side circuit repeater section.

around a general trend line. When a measurement of return loss is made at a single pre-selected value, the return loss obtained is determined by what may be considered as two components, (1) the general quality of the repeater section as determined by the trend line, and (2) the particular part of the " wabble " on which the particular frequency measurement happens to fall (i.e., the bottom or top, or in between point of a given " wabble "). In measuring the return losses of a large group of lines at, say, 2900 cycles, the lower values of return loss will tend to be those which happened to be measured at the bottom of a " wabble." The higher the return loss of a given circuit, the more chance there is that the return loss of that line happened to be measured at the top of a " wabble."

When singing points [4] are measured, however, it may be said that when singing takes place above, say, 2000 cycles, very nearly the minimum value of return loss minus gain is obtained. The distribution of singing points on a large group of lines, therefore, will be lower (in decibels after correction for gain characteristic) for a given percentage than the distribution of single-frequency return losses even at the frequency which has the lowest computed values of any within the transmitted band (say, 2900 cycles for a repeater cutting off about there). Following the reasoning concerning the " wabble " given above, the amount by which the singing point will be lower will be small for the lower return losses and singing points of a given group of lines, and will tend to increase more and more as the value of return loss becomes higher.

Figure 6 shows the S_F curve plotted on a scale which makes all cumulative normal law distributions come out as straight lines. It may be seen that the differences just described will tend to make the singing point distribution more nearly a normal law than the S_F curve. Field measurements show that such singing point measurements do approximate a normal law much more closely than they approximate the S_F curve. Measurements of singing points on about 900 19-gauge H–172–63 side circuits at 18 places during completion tests gave standard deviations about the average of the group at each place which, when added together as the weighted root mean square, gave a general standard deviation of 2.02 db. About 400 similar measurements on 16-gauge H–44–25 side circuits at 16 points gave a standard deviation of 2.05 db. Similar measurements on 233 19-gauge H–88–50 side circuits and 77 19-gauge H–88–50 phantom circuits gave standard deviations of 2.13 and 2.03 db, respectively. It therefore seems reasonable to conclude that a standard deviation of about 2 db is substantially correct for the distribution curve of singing points at a given place. It should be realized that if singing points for a given type of facility from a large number of places are grouped together and a standard deviation of the entire group obtained around the average of the entire group, it may be considerably larger than 2 db due to the differences in average values of the different groups. Such computations on about 7500 measurements on one type of facilities showed a standard deviation of about 3 db for the entire group.

[4] The singing point of a given line is the gain which must be connected between the two sides of a hybrid coil to just cause singing, when the line terminals of the hybrid coil are connected to the line in question and the network terminals are connected to the normal balancing network circuit. Unless otherwise specified, the singing point is expressed in terms of the 1000-cycle gain.

The other important point in determining the singing point distribution is to fix the amount by which its average is less than say, the reference 63 per cent point on the S_F curve; i.e., if the distribution

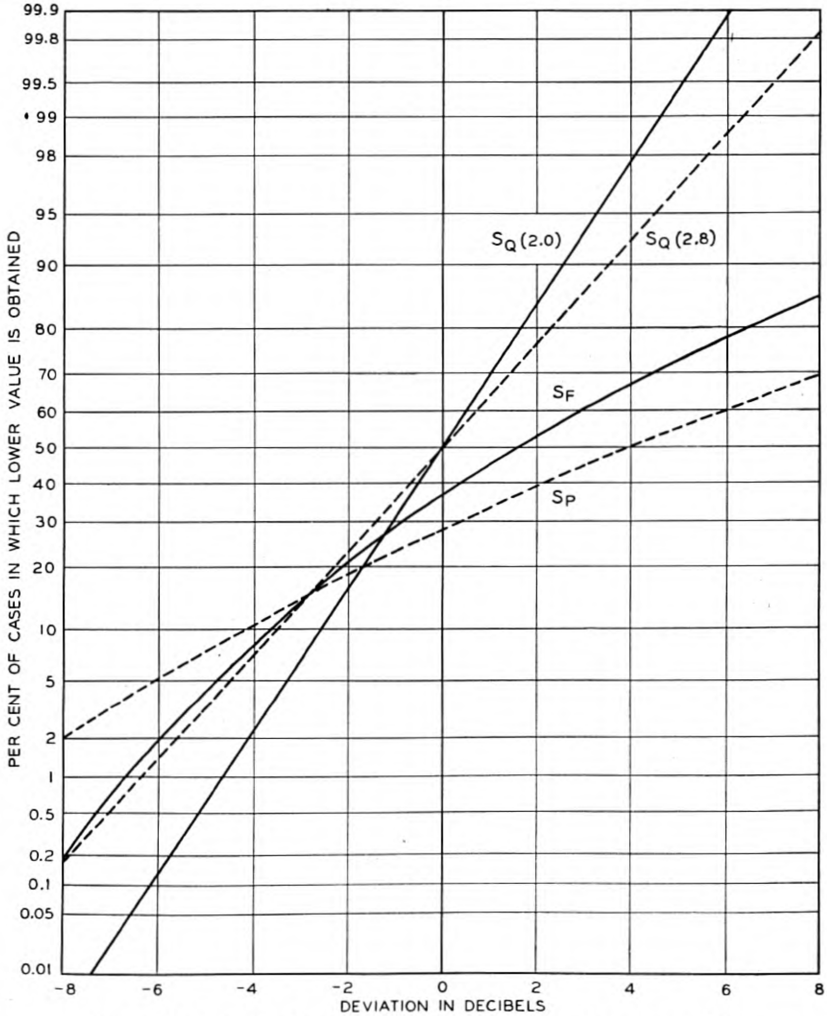


Fig. 6—Distribution functions for return losses and singing points.

curve of single frequency return losses for, say, the highest frequency having as much gain in the repeater as the gain at 1000 cycles, is $S_1 + S_F$, the distribution of singing points will be $S_1 - B + S_Q$ (2.0) where B is S_1 minus the average of the singing points, S_Q (2.0) is a normal law with a standard deviation of 2 db, and Q is the percentage of cases which will have lower singing point values. It was not possible to compute B from the measurements referred to above due to incomplete information.

From field measurements including a large group on 19-gauge B & H–88–50 facilities during a Newark-Philadelphia trial, however, it appears that B is about 2.5 db. As further evidence, the reasoning given above concerning the " wabble " in the return loss distribution when considered together with the standard deviation of 2 db leads to the conclusion that the value of B must lie within a range from about 2 to say, 3 or 4 db for all facilities.

It is accordingly considered reasonable to assume that the distribution of singing points on passive repeater sections at a given place will be a normal law of the form $S_1 - 2.5 + S_Q(2.0)$.[5] The S_Q (2.0) law is shown on Fig. 6.

When the singing points of several such repeater sections in tandem are measured, it seems reasonable to say that the distribution of singing points will be a normal law of the form $S_1 - F(N, T) - 2.5 + S_Q(2.0)$, by analogy with the distribution functions of single-frequency return losses.

When an end path is measured by itself, the singing point in the 2000–3000-cycle range will generally be a little less than would be indicated by single-frequency return-loss measurements. However, the difference between the two sets of measurements will probably be less on the average than the 2.5 db indicated above, since the " wabbles " in measurements of terminal return losses are usually much less than in measurements of repeater section return loss, particularly for the lower terminal return losses which generally are on non-loaded loops. It is accordingly estimated that the distribution of singing points of the end paths is

$$S_{21} - 1 + S_Q(2.0) = 6 + E_1 + S_Q(2.0). \qquad (8)$$

When an end path through a terminal return loss is added to the intermediate paths, tests have shown that the singing point of the resultant will in each case be approximately as if the currents of the singing points of the intermediate paths and the end path added directly. A group of such measurements is shown in Appendix III. The singing point of the resultant will therefore have a distribution of

[5] It is interesting to note here that in an article in *Electrical Communication* for July, 1934, entitled "The Prediction of Probable Singing Points on Loaded Cable Circuits," the law is given as effectively $S_1 - 2.5 + S_Q(1.75)$. The difference between these equations is generally within the limits of experimental error. It may be a real difference, however, since there are certain differences in adjustments of building-out condensers which might be expected to affect the standard deviation.

approximately

$$(S_{11} - 2.5) \underset{i}{X} (S_{21} - 1) + S_Q(2.0), \tag{9}$$

where $\underset{i}{X}$ means that quantities so connected are combined as if their currents added directly, e.g., $S_{11} \underset{i}{X} S_{21} = S_{31}$ means that

$$10^{-S_{11}/20} + 10^{-S_{21}/20} = 10^{-S_{31}/20}.$$

When two end paths are connected in series, say around a four-wire cable circuit, it is believed that singing points in the 2000–3000-cycle range will add directly, because (1) due to the large amount of phase shift around the loop, there will be a large number of frequencies which will have the proper phase shift to permit singing and, (2) the return loss in the frequency range of interest on a given line will change slowly with frequency, which will cause the gain required to produce singing to change only slowly with frequency. On this assumption and assuming the normal law distribution given above, it follows directly from the mathematics of such functions that the distribution of the singing points of two such end paths in series is

$$12 + E_1 + E_2 + S_Q(2.8) = S_{21} - 1 + S_{22} - 1 + S_Q(2.8). \tag{10}$$

On the other hand, if only intermediate paths were present on the two sides of a two-wire repeater, the singing points in the two directions would not, in general, add directly, because they would generally be at different frequencies and when singing occurs at one of these frequencies (or at some new frequency) the sum of the return losses will generally be greater than would be indicated by the sum of the singing points. Tests show that the internal singing margin (the 22-test value corrected for frequency characteristic of the repeater, with only intermediate paths) is about 2.5 db higher on the average than would be indicated by the sum of singing points in the two directions. Similar tests show that when the end path is added so that it is very important compared with the intermediate paths, this average difference becomes about zero.

Also, the average one-way singing point on intermediate paths is about $B = 2.5$ db below the reference single-frequency return loss, as outlined above. Considering the internal singing margin as effectively a 21-test [6] measurement through two return losses in series,

[6] A 21-test is a singing test made by increasing the gain of a repeater until singing occurs, with a line and network connected to the line and network terminals, respectively, of one hybrid coil and a fixed known return loss connected to the other hybrid coil.

it seems reasonable to expect this singing point to be about 2.5 db on the average below the reference single-frequency return loss of the two in series.

From either or both of the above two paragraphs, it may be assumed that the internal singing margin is about

$$M_I = S_1 - F(N_1, T) - G_C + S_1 - F(N_2, T) - g_C - 2.5 + S_Q(2.8). \quad (11)$$

The distribution function $S_Q(2.8)$ is an assumption which seems reasonable from general considerations.

From the above discussion the singing margin of the entire circuit will be a function of all the intermediate and end paths which meets the following conditions:

1. When the end paths are unimportant compared to the intermediate paths, the singing margin is

$$M_I = S_{11} + S_{12} - 2.5 + S_Q(2.8). \quad (12)$$

2. When the intermediate paths are unimportant compared to the end paths, the singing margin is

$$M_E = S_{21} - 1 + S_{22} - 1 + S_Q(2.8). \quad (13)$$

3. When both kinds of paths are fairly important, the singing margin is some compromise between

$$(S_{11}) \underset{p}{\mathrm{X}} (S_{21}) + (S_{11}) \underset{p}{\mathrm{X}} (S_{21}) + S_Q(2.8),$$

which would have an average value equal to the reference value of the circulating current margin and about 2.5 db less than this. The singing margin is obviously greater than

$$(S_{11} - 2.5) \underset{i}{\mathrm{X}} (S_{21} - 1) + (S_{12} - 2.5) \underset{i}{\mathrm{X}} (S_{22} - 1) + S_Q(2.8)$$

because the active singing point in one direction is almost certain to be at a different frequency from that of the active singing point in the other direction.

A reasonable empirical compromise which satisfies all of these conditions and preserves the symmetry of the equation is to say that the total singing margin of the circuit is

$$M_s = (S_{11} - 1.25) \underset{p}{\mathrm{X}} (S_{21} - 1) + (S_{12} - 1.25) \underset{p}{\mathrm{X}} (S_{22} - 1) + S_Q(2.8). \quad (14)$$

This may be written in more general fashion as

$$M_s = (S_1 - F(N_1, T) - G_C - 1.25) \underset{p}{X} (6 + E_1) + (S_1 - F(N_2, T) - g_C - 1.25) \underset{p}{X} (6 + E_2) + S_Q(2.8) \quad (15)$$

or still more generally,

$$M_s = (S_H + S_w - S_A - L + T + q - F(N_1, T + q) - 1.25) \underset{p}{X} (6 + E_1' + 2N_1q) + (S_H + S_w - S_A - L + T + q - F(N_2, T + q) - 1.25) \underset{p}{X} (6 + E_2' + 2N_2q) + S_Q(2.8). \quad (16)$$

Criterion of Satisfactory Performance

The above discussion derives methods of determining the circulating current margin and the singing margin that will be obtained (on a distributional basis) for a cable circuit under a given set of conditions. As a practical matter, the overall net loss of the circuit will sometimes vary, regulating repeaters will change in gain setting, temporary troubles will occur and, in some instances, toll circuit terminations will be removed while connections are being set up. Each of these factors will reduce the singing margin that the circuit had under average conditions. In the Bell System toll circuits are usually designed so that a 10 db singing margin or more is obtained in 90 per cent of the loaded two-wire cable circuits under average conditions. This, of course, is the same thing as saying that 12 db singing margin should be exceeded under average conditions in 71 per cent of the cases, or 8 db in 97.7 per cent of the cases (see $S_Q(2.8)$ curve on Fig. 6). The following table shows the percentages of circuits which will have various lower singing margins under average conditions than the indicated values, for various different assumptions as to the design requirement (i.e., the singing margin which must be exceeded in 90 per cent of the cases).

Design Requirement in Db	Singing Margin in Db =	Percentages Having Lower Singing Margins					
		14	12	10	8	6	4
12		29	10	2.3	0.23	0.03	0.002
10		57	29	10	2.3	0.33	0.03
8		84	57	29	10	2.3	0.33
6		95	84	57	29	10	2.3

The following table shows the same information for circulating current margins, computed on the assumption that $S_{11} - 1.25 = S_{12} - 1.25 = (S_{21} - 1) = (S_{22} - 1)$. The design requirements in this case are also that 12 db, 10 db, 8 db or 6 db singing margin (not circulating current margin) shall be exceeded in 90 per cent of the cases under average conditions.

Design Requirement in Db	Percentages Having Lower Circulating Current Margins					
Circulating Current Margin in Db =	14	12	10	8	6	4
12	11.3	5.55	2.25	0.78	—	—
10	19.4	11.3	5.55	2.25	0.78	—
8	29.4	19.4	11.3	5.55	2.25	0.78
6	40.2	29.4	19.4	11.3	5.55	2.25

It is estimated that the reduction in circulating current margin due to regulation, variation in circuit net loss, and temporary troubles will seldom exceed $4\frac{1}{2}$ db by very much but about this amount of reduction will occur fairly frequently. Similarly, the reduction in singing margin due to these causes and also to the removal of a termination while a circuit is being set up will seldom exceed $5\frac{1}{2}$ db. It is believed that about 4 db circulating current margin at the critical frequency is as small as can be obtained without very objectionable quality distortion.

When $5\frac{1}{2}$ db reduction in singing margin is obtained on the average, we may read from the first table above at $5\frac{1}{2}$ db margin, the percentage of cases in which actual singing will take place. For example, with a design requirement of 8 db, interpolating between the vertical columns headed 6 and 4 db, respectively, we find that about $1\frac{1}{2}$ per cent of cases will result in actual singing. Similarly, by reading at $4 + 4\frac{1}{2} = 8\frac{1}{2}$ db circulating current margin in the second table, we find the percentage of cases in which the quality will be objectionably distorted during the period in which a $4\frac{1}{2}$ db reduction in circulating current margin is obtained. The following table shows these two different percentages for the design limits given above. From this table it

Design Limit in Db	Percentage of Cases With Less Than 4 Db Circulating Current Margin Under Extreme Conditions	Percentage of Cases Which Will Sing Under Extreme Conditions
12	0.97	0.015
10	2.8	0.18
8	6.8	1.45
6	13.1	7.1

seems reasonable to require a 10 db singing margin under average conditions in order to keep the percentage of circuits which will have objectionable circulating currents and the percentage which will sing occasionally low enough that serious service reactions will not be obtained.

Practical Considerations

Equipment and Terminating Effects

When actual cable sections are investigated, the effect of the equipment must be considered. The effect of adding, say, repeating coils in the line and network sides of the repeater, is to introduce some reflected current and to add a certain loss in the line which must be made up by increasing the repeater gain. The cable return loss is increased by twice the loss introduced, so that the only effect on circulating current paths is due to the differences in the line and network equipment.

Appendix IV derives an expression for the distribution curve of an active or passive single-frequency return loss of a line with equipment, on certain assumptions. The complication of using this method to determine circulating current margins or singing margins is large, however, and an approximate method is illustrated below.

Figure 7 shows an example of the combination of the equipment-return loss with the return loss of a cable section without equipment. It is assumed in this case that the equipment-return loss is 13 db higher than the value of cable-return loss for which $S_F = 0$. For example, the equipment return loss might be 40 db while the line return loss was $27 + S_F$ db. A generally similar curve would be obtained from the combination of the repeater section terminating effect with the cable return losses.

Since the principal interest in connection with these return losses is with the lower values, it seems reasonable to say that the effect may be approximated generally by shifting the S_F curve a certain part of a decibel, depending upon the difference between S_E and the cable-return losses. Strictly speaking, of course, it not only shifts the curve but also changes the standard deviation, but the complication of taking account of this does not seem worth while with present lines and equipment.

Since we are interested in the singing margin which is exceeded 90 per cent of the time under average conditions, it would be possible to select a value of cable return loss from the distribution curve which, if it occurred in each repeater section of a particular circuit, would give the indicated singing margin when the various intermediate singing paths are combined together as the sum of their power ratios.

This would be roughly where $S_F = -2$ db. The percentage of cable section return losses which would exceed this value, however, would be about 80 per cent, so a line which gave such a low singing margin would, in general, tend to be made up of a few very low-return losses

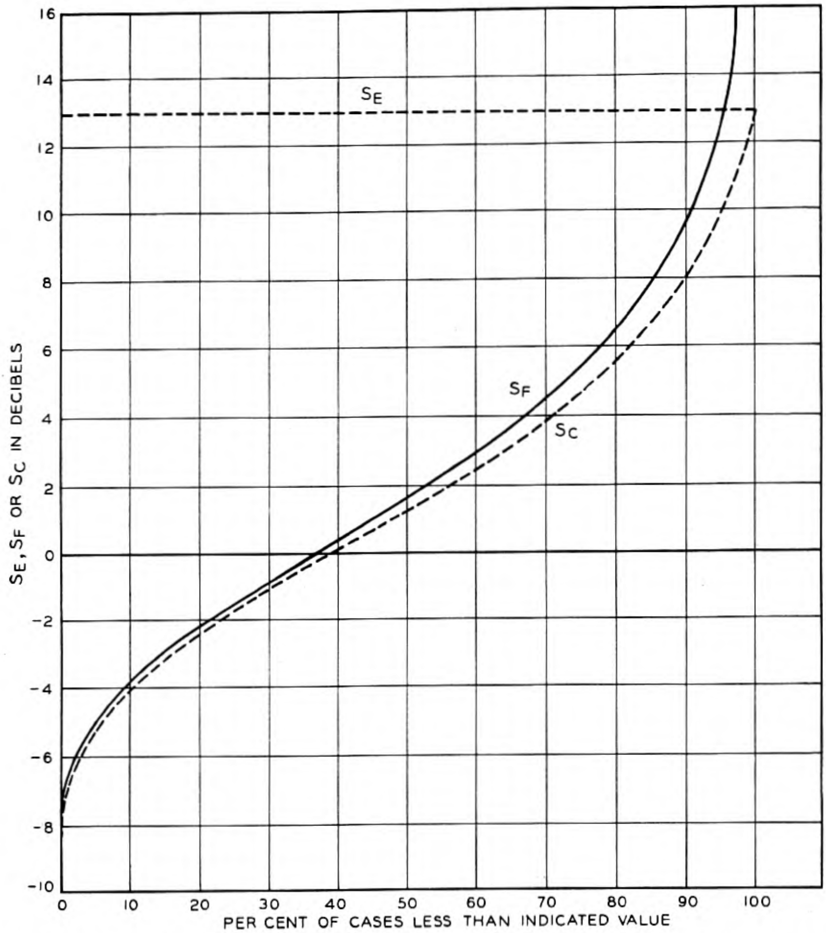


Fig. 7—Return loss distributions of cable pairs (S_F), equipment (S_E), and the combination (S_C).

and a considerable percentage of higher-return losses. It, therefore, seems that this value of cable-return loss, if combined with the equipment and terminating return loss, would not be sufficiently degraded to consider that degradation as a fair average degradation for the entire curve. On the other hand, if the cable return loss which is exceeded in 50 per cent of the cases is taken (i.e., $S_F = + 1.8$ db), the degradation would evidently be too large due to the equipment since the lines in which the singing margin might be expected to be low, would generally be those which happen to have more than the usual number of low line return losses.

As an approximation in the usual case, it seems reasonable to combine the line equipment return losses with the cable return losses as the sum of their power ratios, with a value of S_F of about $- 1$ db (i.e., where 28 per cent of the return losses are lower than the value considered). One db added to the resultant will give the value of the resultant for which $S_F = 0$. The same thing may also be accomplished by combining the cable return loss for $S_F = 0$ with the near and far-end equipment return losses each increased by 1 db. The example which is given later shows the method used.

Taper

The taper of a regular two-wire circuit is the amount of decrease in repeater output level at a given frequency which occurs between succeeding repeaters from the transmitting to the receiving end of the circuit. For example, starting at the transmitting repeater, if the repeater output levels at 1000 cycles were, respectively, 3.0, 2.5, 2.0, 1.5, etc., the 1000-cycle taper would be 0.5 db.

An optimum taper from the standpoint of singing will be reached when the singing margin around the most critical intermediate repeater is equal to the singing margin around the terminal repeater. This is certain to occur for some taper under the present method of distributing gain since an increase of the taper causes the gain of the receiving terminal repeater to be increased by a multiple of the increase of taper. However, the optimum taper from the standpoint of singing is not generally used in the field, because the optimum taper from the standpoints of crosstalk and noise is zero db. A compromise is generally made on two-wire cable circuits; e.g., 19-gauge B and H–88–50 facilities use 0.5 db and 16-gauge H–44–25 facilities use 0.2 db taper.

Active Balances

In some cases, it may be more convenient to measure active return losses or active singing points rather than singing margins. The expected distribution of active return losses and active singing points including the end paths when the end paths are the normal working terminations, have been discussed above.

In practice, however, the termination when such an active singing point test is made will normally be a fixed one, generally consisting of resistances with a return loss substantially the same at all frequencies. Appendix V derives the distribution of active singing points for these cases.

For exact results, the modification of the return loss due to the near-end and far-end equipment should be added before combining the end path. In the practical case, however, only the lower values of return loss are generally significant, and the approximate method outlined above may generally be used without serious error.

Minimum Value of a Group of Singing Points

When a large number of independent measurements are taken, say of singing points on different pairs at a given repeater station, it is often of interest to know whether the maximum or minimum in that group represents some special condition, say, definite circuit trouble, or is merely a case where the various components happen to have added up or subtracted by pure chance.

In the case of singing points, the question of interest may usually be stated as follows: In " n " similar measurements at a given place, what is the probability that the smallest value will be more than x db below the true average measurement, assumed known?

If the fraction of cases which will be greater than a value y db below the average is P, the probability that " n " values will all be greater than this value is P^n. In other words $P_m(n, y) = P_m = P^n$ is the probability that no one in " n " values will be as much as y db below the true average.

For example, if the singing point distribution is $25 + S_Q(2)$, the chance that any given value selected at random will be as low as $25 - 4.6 = 20.4$ db is one in a hundred. However, the chance that the lowest one of 20 measurements will be that low is $1 - (0.99)^{20} = 0.1819$ or almost one in five. If we substitute $V_m = 1 - P_m$ and $V = 1 - P$, we may write that $1 - V_m = (1 - V)^n$.

$$V_m = 1 - (1 - V)^n = 1 - 1 + nV - \frac{n(n-1)}{2!} V^2 + \cdots$$

$$= nV - \frac{n(n-1)V^2}{2!} + \frac{n(n-1)(n-2)}{3!} V^3 - \cdots.$$

And for small values of V, $V_m = nV$. In the above case, therefore, if we wanted to compute a value of singing point which would be lower than all the 20 measurements nine times out of ten, we may compute $V = (0.1/20) = 0.005$, and read on the normal distribution $S_Q(2)$ that this corresponds to about 5 db below the average; i.e., to $25 - 5 = 20$ db. The more exact formula would be that $0.9 = P^{20}$

or 20 log P = log 0.9 or log $P = -0.002288$ or $P = 0.99476$ or $V = 0.00524$. Reading on the normal law distribution for this value changes the 20 db limit as computed above to about 20.03 db, the difference being negligible compared to the usual accuracy of such measurements.

In general, measurements outside of the limits so determined are probably cases of trouble. However, in the case of active singing points, certain systematic changes in the measured values from the computed values must be allowed for before trouble is proved.

Example

Following is an example of the application of these various methods.

Given: The following losses at the critical frequency for various successive repeater sections shown in Fig. 1.

Repeater Section	Loss in Db
A to B	17
B to C	17
C to D	18
D to E	15
E to F	15
F to G	16
Total	98

The following gains and return losses at the critical frequency are assumed. The gains give a 9 db overall net loss at the critical frequency.

At Point	Repeater Gain at the Critical Frequency—db			Return Losses at the Critical Frequency for $S_F = 0$, i.e., Values Exceeded in 63 Per Cent of Cases—db							
				Bare Line **		Near-End Apparatus *		Far-End Apparatus †		Combination ‡	
	A to G	G to A	Sum	Toward G	Toward A	Toward G	Toward A	Toward G	Toward A	Toward G	Toward A
A	2.4	12.0	14.4	27.6	—	41	—	46	—	27.35	—
B	15.0	15.0	30.0	27.6	27.6	41	41	46	46	27.35	27.35
C	16.8	16.9	33.7	27.6	27.6	41	41	48	46	27.37	27.35
D	16.9	13.9	30.8	27.6	27.6	41	41	42	48	27.25	27.37
E	13.9	13.9	27.8	27.6	27.6	41	41	42	42	27.25	27.25
F	13.9	14.9	28.8	27.6	27.6	41	41	44	42	27.32	27.25
G	10.1	2.4	12.5	—	27.6	—	41	—	44	—	27.32
Total	89.0	89.0	178.0								

* This is 40 db nominal apparatus return loss plus 1 db allowance as described above.

† This is 11 db far-end equipment terminating return loss plus 1 db allowance as described above plus twice the loss from the measuring point to the far-end equipment.

‡ These are the combinations of the three preceding return losses as the sums of their power ratios.

** Twice the loss of the near-end equipment is included.

Considering C as the critical repeater:

Path		
Toward A	1	27.35
	2	$27.35 + 34 - 30 = 31.35$
	Total	25.90 Items 1 and 2 combined as the sum of their power ratios.
Toward G	1	27.37
	2	$27.25 + 36 - 30.8 = 32.45$
	3	$27.25 + 38 - 31.8 + 30 - 27.8 = 35.65$
	4	$27.32 + 38 - 31.8 + 30 - 27.8 + 30 - 28.8 = 36.92$
	Total	25.42. Items 1, 2, 3 and 4 combined as the sum of their power ratios.

The internal circulating current margin then is $25.90 + 25.42 - 33.7 = 17.62$ db, or more in 72 per cent of the cases and the internal singing margin is $25.90 + 25.42 - 2.5 - 33.7 = 15.12$ db or more in 50 per cent of the cases, both under the average conditions specified.

Now 10 db internal circulating current margin would be exceeded under average conditions in the percentage of cases for which S_P is $10 - 17.62 = -7.62$ db or greater; i.e., in about 97.5 per cent of the cases, reading from the S_p curve on Fig. 6. Or, similarly, 10 db internal singing margin will be exceeded in about 97 per cent of the cases, reading at $S_Q(2.8) = 10 - 15.12 = -5.12$ from the $S_Q(2.8)$ curve on Fig. 6.

The single-frequency loss in the end path toward A, before the gain of repeater C is introduced is, for a net loss at the critical frequency of 9 db, $34 + 34 - 30.0 - 14.4 + 7 + S_F = 30.6 + S_F$. In the other direction the corresponding function is $36.1 + S_F$. When the net loss of the circuit is expressed as E rather than 9 db, these functions become respectively $21.6 + E + S_F$ and $27.1 + E + S_F$. The corresponding singing point distributions are $20.6 + E + S_Q(2)$ and $26.1 + E + S_Q(2)$, respectively.

The desired value of 10 db or more singing margin obtained in 90 per cent of the cases under average conditions will be obtained when the following equation is satisfied:

$$10 = (25.90 - 1.25) \underset{p}{\mathrm{X}} (20.6 + E) + (25.42 - 1.25) \underset{p}{\mathrm{X}} (26.1 + E) - 33.7 + S_Q(2.8), \quad (17)$$

where $Q = 10$ per cent, $S_Q(2.8) = -3.5$ db. Substituting in the equation and simplifying,

$$- 1.62 = 0 \underset{p}{X} (E - 4.05) + 0 \underset{p}{X} (E + 1.93), \tag{18}$$

which is satisfied if $E = 8.8$ db.

The active balance of the circuit with a 9 db net loss at the critical frequency measured from the " A " end may be computed as follows: The successive paths from the line side of repeater " A " are (using values for $S_F = 0$):

1. 27.35
2. $27.35 + 34 - 30 = 31.35$
3. $27.37 + 34 - 30 + 34 - 33.7 = 31.67$
4. $27.25 + 34 - 30 + 34 - 33.7 + 36 - 30.8 = 36.75$
5. $27.25 + 34 - 30 + 34 - 33.7 + 36 - 30.8 + 30 - 27.8 = 38.95$
6. $27.32 + 34 - 30 + 34 - 33.7 + 36 - 30.8$
   $+ 30 - 27.8 + 30 - 28.8 = 40.22.$

Combining these paths as the sum of their power ratios gives 24.33 db, which is the computed single-frequency return loss without the end path from the line side of the terminal repeater for $S_F = 0$ (i.e., which will be exceeded in 63 per cent of the cases). The active singing point from this point will be $24.33 - 2.5 + S_Q(2) = 21.83 + S_Q(2)$. The corresponding active singing point referred to the drop side of the repeater will be $21.83 - 14.4 + S_Q(2) = 7.43 + S_Q(2)$ db. With a fixed termination giving a return loss of 5 db at the circuit terminal, the end path from the drop side of the repeater will be $18 + 5 = 23$ db. From Appendix V, for $d = 23 - 7.43 = 15.57$ db, the expected active singing point including the end path is the distribution curve 7.43 $+ S_g(15.57)$ under average conditions. For example, by interpolation in the table in Appendix V, two per cent of such circuits will have lower active singing points from the drop than about $7.43 - 5 = 2.43$ db and from the line side of the repeater than $2.43 + 14.4 = 16.83$ db, both under average conditions.

This answer may be computed by reading from the $S_Q(2.0)$ at two per cent which gives $S_2(2.0) = - 4.1$ db. At this percentage, therefore, the active singing point without the end path is $7.43 - 4.1 = 3.33$ db. Combining this with the fixed and path of 23 db gives $3.33 - 0.9 = 2.43$ db.

APPENDIX I

Active Return Loss Of Intermediate Paths

Let L_k = the loss from the west side of repeater C to the output of repeater X which is the k th repeater to the west, plus the loss from the input of the W–E side of repeater X back to the input of repeater C (Fig. 1).

$$\text{Let } L_k = 20 \log_{10} t_k^{2k}. \tag{19}$$

Now from Equations (26) and (27) from " Irregularities in Loaded Telephone Circuits ": [7]

$$I_{1k} = \frac{I_0 R_L}{t_k^{2k}} \sqrt{\sum_{j=1}^{m} A_k^{4(j-1)} \cos^2 \theta_{jk}}, \tag{20}$$

$$I_{2k} = \frac{I_0 R_L}{t_k^{2k}} \sqrt{\sum_{j=1}^{m} A_k^{4(j-1)} \sin^2 \theta_{jk}}, \tag{21}$$

where I_{1k} is I' (from the paper) for the currents from the $(k + 1)$ th repeater section from C (counting the adjacent repeater section as the first) and I_2k is I'' for the same section.

The total components will be

$$I_1 = I_0 R_L \sqrt{\sum_{j=1}^{m_k} \sum_{k=0}^{N-1} \frac{A_k^{4(j-1)}}{t_k^{4k}} \cos^2 \theta_{jk}}, \tag{22}$$

$$I_2 = I_0 R_L \sqrt{\sum_{j=1}^{m_k} \sum_{k=0}^{N-1} \frac{A_k^{4(j-1)}}{t_k^{4k}} \sin^2 \theta_{jk}}, \tag{23}$$

where N = the total number of repeater sections so considered.

Assuming (as in the paper) that $I_1 = I_2$

$$I_1 = I_2 = \frac{I_0 R_L}{\sqrt{2}} \sqrt{\sum_{k=0}^{N-1} \frac{(1 - A_k^{4m_k})}{t_k^{4k}(1 - A_k^{4})}}. \tag{24}$$

By analogy, Equation (34) from the paper referred to may be used for the distribution of the total current I_F, with $I_1 = I'$; i.e.,

$$F = e^{-(I_F^2/2I_1^2)}. \tag{25}$$

Equation (42) from the paper may therefore be written

$$S = S_H + S_W + S_F - S_A - F(N, T) \tag{26}$$

where

$$S_A + F(N, T) = 10 \log_{10} \left(\sum_{k=0}^{N-1} \frac{(1 - A_k^{4m})}{t_k^{4k}(1 - A_k^{4})} \right). \tag{27}$$

[7] G. Crisson, October 1925, *Bell Sys. Tech. Jour.*

Where all values of A_k are the same, and equal to A,

$$S_A = 10 \log_{10}\left(\frac{1 - A^{4m}}{1 - A^4}\right) \tag{28}$$

$$F(N, T) = 10 \log_{10}\left(\sum_{k=0}^{N-1} t_k{}^{-4k}\right). \tag{29}$$

Where all values of " t_k " are the same and equal to " t,"

$$F(N, T) = 10 \log_{10}\left(\frac{1 - t^{-4N}}{1 - t^{-4}}\right). \tag{30}$$

This is the case where all the gains are equal and all the losses are equal at all points.

In the case where $t = 1$, i.e., $T = 0$ db,

$$F(N, O) = 10_{10} \log N. \tag{31}$$

APPENDIX II

Distribution of Two Active Return Losses in Series

Assume two return losses in series, the two being selected at random from groups of return losses (active or passive) having distributions represented by $S_{11} + S_F$ and $S_{12} + S_F$, respectively. What is the distribution curve of the two in series?

Let

$$p_1 = \frac{i_1}{k_1{}^2} e^{-i_1{}^2/2k_1{}^2}, \tag{32}$$

$$p_2 = \frac{i_2}{k_2{}^2} e^{-i_2{}^2/2k_2{}^2}, \tag{33}$$

where p_1 is the relative probability of obtaining a returned current i_1 from the first line ($S_{11} + S_F$) and k_1 corresponds to I_1 in Appendix I. The quantities with subscript 2 are the same for the second line ($S_{12} + S_F$). Now if $i_1 i_2 = i_3$, the probability that i_3 exceeds a certain value I is

$$P(i_3 > I) = \int_{i_2=0}^{\infty} \int_{i_1=I/i_2}^{\infty} p_1 di_1 p_2 di_2. \tag{34}$$

Substituting from Equations (32) and (33) and letting

$$x = \frac{i_2{}^2}{2k_2{}^2}, \qquad a = \frac{I}{2k_1k_2}, \tag{35}$$

$$P(1_3 > I) = \int_0^{\infty} e^{-(x+(a^2/x))} dx, \tag{36}$$

$$P(1_3 > I) = 2aK_1(2a), \tag{37}$$

where K_1 is a Bessel's function of the second kind (see " Theory of Bessel Functions " by G. N. Watson).

Now

$$a = \frac{I}{2k_1k_2} = \text{anti-log}_{10}\left(\frac{S_{11} + S_{12} - S_I}{20}\right), \tag{38}$$

where S_I is the return loss obtained from the two in series.

Let

$$S_p = -20 \log_{10} a = -S_{11} - S_{12} + S_I, \tag{39}$$

$$S_I = S_{11} + S_{12} + S_p. \tag{40}$$

This last expression is the distribution curve of the two return losses in series.

APPENDIX III

Addition of End Path to Active Singing Points
(16 Ca. H-44-25 2-Wire Cable Side Circuits)

(1) Number of Repeater Sections	(2) Active Singing Point to Terminal Repeater on Zero—db	(3) Loss of End Path—db	(4) Active Singing Point with End Path Added—db	(5) (2) and (3) Added as Sum of Current Ratios	(6) (2) and (3) Added as Sum of Power Ratios
4	29.8	20.4	17.1	17.9	19.0
4	31.2	20.4	18.5	18.2	20.2
4	32.3	23.8	21.3	21.0	23.2
4	31.2	23.8	21.2	20.7	23.1
4	33.1	23.8	20.9	21.2	23.3
4	29.3	20.8	17.8	18.0	20.2
4	35.7	20.8	18.7	19.3	20.7
4	27.0	20.8	18.0	17.3	19.9
8	25.6	22.2	19.0	17.7	21.6
8	24.2	19.9	16.1	15.8	18.5
18	17.9	19.1	12.7	12.4	15.4
18	17.7	19.1	13.5	12.4	15.3
18	20.1	25.2	15.5	16.2	18.9
18	18.7	25.2	15.1	15.3	17.8
18	20.0	19.0	14.5	13.5	16.5
18	16.9	19.0	12.0	11.9	14.8
18	19.1	25.1	16.2	15.6	18.1
18	19.9	25.1	17.2	16.1	18.8
18	21.0	20.9	14.0	14.9	17.9
18	20.1	24.3	14.1	14.5	17.5
18	19.0	24.3	15.7	15.2	17.9
18	17.8	24.3	15.0	14.4	16.9
		Average =	16.45	16.34	18.88

The different tests using the same number of repeaters are different circuits and different testing repeaters. Columns (2) and (4) are measured singing points. Column (3) is the sum of the measured losses from the measuring point to the drop side of the terminal repeater and back, plus the return loss of resistance termination applied at the drop side of the repeater.

Columns (5) and (6) show computed values corresponding to column (4), on different assumptions as to the method of addition of the intermediate and end paths. The values in column (5) are generally much closer to the values in column (4) than are those in column (6).

APPENDIX IV

Distribution of Active Return Losses with Intermediate Paths Including Equipment

From a consideration of the paper, "Irregularities in Loaded Telephone Circuits," by G. Crisson,[7] it appears that it is reasonable to say that the total current returned from a cable section with near-end equipment will be

$$i_t = \sqrt{i_1^2 + i_e^2},$$

where i_1 is the current returned from the line without equipment and i_e is the current returned from the equipment. (It is assumed here that i_t and i_e are referred to the line side of the coil and not the drop side.)

Then the probability that the current returned from the cable alone will exceed I is

$$F(I) = \frac{1}{I_1^2}\int_I^\infty i_1 e - \frac{i_1^2}{2I_1^2} di_1 = e - \frac{I^2}{2I_1^2}.$$

The probability that the total current will exceed $I_3 = \sqrt{I^2 + i_e^2}$ is

$$P(i_t > I_3) = e - \frac{(I_3^2 - i_e^2)}{2I_1^2} = e \frac{i_e^2}{2I_1^2} F(I_3)$$

or

$$P(i_t > I_3) = \text{antilog}_{10}\left(.434 \text{ antilog}_{10}\left(\frac{S_1 - S_E}{10}\right)\right) F(I_3) = F',$$

where

$$S_1 = -10 \log_{10} 2I_1^2 \qquad \text{and} \qquad S_E = -20 \log_{10} i_e.$$

This equation says that the probability of getting more than a given value of return current I_3 from the combined cable section and near end equipment is equal to the product of the probability of getting I_3 from the cable section alone by the factor $e(i_e^2/2I_1^2)$. Or by analogy with the method used in the paper referred to above, letting the distribution curve of the combination return loss be $S_1 + S_{F'}$,

$$S_{F'} = -\log_{10}\left(\log_e \frac{1}{F} + \frac{i_e^2}{2I_1^2}\right)$$

$$= -\log_{10}\left(\log_e \frac{1}{F} + \text{antilog}_{10} \frac{S_1 - S_E}{10}\right)$$

$$S_{F'} = S_F \underset{p}{\text{X}} (S_E - S_1).$$

[7] Loc. cit.

Or the distribution curve of the return losses from the repeater section with equipment is

$$S_1 + S_{F'} = (S_1 + S_F) \underset{p}{\text{X}} (S_1 + S_E - S_1)$$

$$= (S_1 + S_F) \underset{p}{\text{X}} (S_E)$$

$$S_{F'} = (S_F) \underset{p}{\text{X}} (S_E - S_1).$$

Curves of $S_{F'}$ could be drawn for different values of $S_E - S_1$, since S_F is known for a given value of F. These would apply either for passive return losses of a single repeater section or for active return losses of several repeater sections in tandem.

For example, suppose $S_E = S_1$. For $S_F = 0$, the value of F is about 0.368 and the value of $S_{F'} = 0 \underset{p}{\text{X}} 0 = -3.01$ db for $F' = 0.368$. Or the same point could be derived from the earlier formulas as follows: When $S_F = -3.01$ db, $F = 0.1353$. The factor antilog $\left(0.434 \text{ antilog}\left(\frac{0}{10}\right)\right)$ equals 2.718 and the value of F' for $S_{F'} = -3.01$ db is $F' = (0.1353)(2.718) = 0.368$.

APPENDIX V

Distribution of Active Return Loss Made up of Intermediate Paths and a Fixed End Path

In a practical case, when an active singing point toward a circuit terminal is measured for maintenance purposes, the termination at the circuit drop will be a fixed known value of return loss rather than one selected at random from a distribution curve. The distribution curve of active singing points toward the drop will then be a curve obtained at each percentage, by the current sum addition of the active singing point with no current returned from beyond the terminal repeater and the end path singing point. If the active intermediate path singing point has a distribution curve $S_{11} - 2.5 + S_Q(2) = S_4 + S_Q(2)$ and the end path has a fixed loss S_5, the following table may be used to find the distribution curve of the combination $S_4 + S_g(d)$, for different values of $d = S_5 - S_4$. The table was derived from the equation

$$(S_4 + S_Q(2)) \underset{p}{\text{X}} S_5 = S_4 + S_g(d)$$

Percentage of Cases with Lower Singing Points	0.003	0.13	2.2	15.7	50.0	84.3	97.8	99.87	99.997	99.99997
$S_g(\infty)$ in db..	−8	−6	−4	−2	0	2	4	6	8	10
$S_g(15)$ in db..	−8.6	−6.7	−4.9	−3.2	−1.4	+0.2	+1.8	+3.4	+4.8	+6.1
$S_g(10)$ in db..	−9.0	−7.3	−5.6	−4.0	−2.4	−0.9	+0.5	+1.1	+2.9	+4.0
$S_g(6)$ in db...	−9.6	−8.0	−6.4	−4.9	−3.5	−2.2	−1.1	0	+0.9	+1.8
$S_g(3)$ in db...	−10.2	−8.6	−7.2	−5.9	−4.6	−3.5	−2.5	−1.6	−0.9	−0.2
$S_g(0)$ in db...	−10.9	−9.5	−8.2	−7.1	−6.0	−5.1	−4.2	−3.5	−2.9	−2.4
$S_g(-3)$ in db.	−11.9	−10.6	−9.5	−8.5	−7.6	−6.9	−6.2	−5.6	−5.2	−4.8
$S_g(-6)$ in db.	−13.1	−12.0	−11.1	−10.2	−9.5	−8.9	−8.4	−8.0	−7.6	−7.3
$S_g(-10)$ in db	−15.1	−14.2	−13.5	−12.9	−12.4	−12.0	−11.5	−11.3	−11.0	−10.8

or

$$(S_Q(2)) \underset{p}{\mathrm{X}} (S_5 - S_4) = S_g(d).$$

It will be noted that $S_g(\infty) = S_Q(2.0)$.

It should be noted that the values of $S_g(d)$ corresponding to the higher percentages may be considerably modified by the effect of the equipment return losses, particularly when d is large. However, this is usually of no very great importance because it is the lower percentages and the smaller values of d which are of the most practical importance.

APPENDIX VI

Nomenclature

A = attenuation factor per loading section (current ratio).

A_k = attenuation factor per loading section (current ratio) of k th repeater section from critical repeater.

a $= \log_{10}^{-1}\left(\dfrac{S_{11} + S_{12} - S_I}{20}\right)$.

B = S_1 minus the average singing point on a group of lines.

b = the standard deviation of a group of singing point measurements.

d $= S_5 - S_4$.

E = net loss of circuit in db.

E_1 = loss in end path toward West for zero terminal return loss. When primed, it is the loss at 1000 cycles.

E_2 = same as E_1 but toward East.

$F(I_F) = F$ = probability of obtaining a greater returned current then I_F from a randomly selected one of a group of lines.

F' = probability of obtaining a greater returned current than I_3 from a cable section with near end equipment.

$F(N, T)$ = $10 \log_{10} \left(\sum_{k=0}^{N-1} t_k^{-4k} \right)$.

G_C and g_C = gains as defined herein.

I = current obtained in a particular case.

I_0 = current sent into the line at the point of measurement.

I_1 = $\text{anti} \log_{10} \left(\frac{- S_1 - 3}{20} \right)$ = current for active return loss corresponding to I' from " Irregularities in Loaded Telephone Circuits " by G. Crisson in the October 1925 *Bell System Technical Journal.*

I_2 = $\text{anti} \log_{10} \left(\frac{- S_1 - 3}{10} \right)$ = current for active return loss corresponding to I'' from paper referred to under I_1.

I_3 = current defined above under F'.

I_F = current defined under $F(I_F)$ above.

I_{1k} = I' (from paper referred to above) for the currents from the k th repeater section from the critical repeater.

I_{2k} = I'' (from paper referred to above) for the currents from the k th repeater section from the critical repeater.

i_1 and i_2 = currents returned from particular active or passive return losses.

i_3 = current returned from a particular pair of active or passive return losses in series.

i_e = $\text{anti} \log_{10} \left(\frac{- S_E}{20} \right)$.

i_t = total current returned from line with equipment.

j = number of the loading section measured from the repeater at which the return loss is of interest.

$K_1(x)$ = Bessel function of the second kind and order unity with argument x.

k = the number of the repeater or repeater section from the critical repeater, e.g., for the adjacent repeater section $k = 1$, for the next $k = 2$, etc.

k_1 $= \text{anti} \log_{10}\left(\frac{-S_{11} - 3}{20}\right)$ = current corresponding to S_{11} in same way as I_1 corresponds to S_1.

k_2 $= \text{anti} \log_{10}\left(\frac{-S_{12} - 3}{20}\right)$ = current corresponding to S_{12} in same way as I_2 corresponds to S_1.

L = loss of a cable section.

L_k = the loss in db from the West side of the critical repeater to the output of the k th repeater to the West, plus the loss from the input of the k th repeater W–E to the input of the critical repeater W–E.

M_E = the end path singing margin, i.e., the singing margin with no currents returned from intermediate paths (e.g., with a four-wire circuit).

M_I = the internal singing margin on an active line, i.e., the singing margin without any currents returned from the end paths (e.g., with the terminal repeaters turned down).

M_S = the singing margin of the circuit; i.e., the singing margin with currents returned both from intermediate paths and from end paths.

m_k = total number of loading sections in the k th repeater section.

m = value of m_k when all repeater sections have the same number of loading sections.

N = number of repeater sections in the part of the circuit for which the active return loss is to be computed.

n = number of cases considered as a group.

p_x = relative probability of obtaining a returned current i_x from one active or passive return loss.

$P(x > y)$ = probability that the value of x selected at random from a distribution curve is greater than a pre-selected value y.

P = probability that a singing point will have a lower value than a pre-selected value y.

$P_m(n, y) = P_m$ = probability that the minimum of n values will be as high or higher than a pre-selected value y db below the true average.

Q = probability that a lower value of singing margin will be obtained.

q	= frequency taper in db per repeater section.
R_L	= the representative reflection coefficient (see paper by G. Crisson referred to above).
S_H	= the irregularity function (see paper by G. Crisson referred to above).
S_W	= the frequency function (see paper by G. Crisson referred to above).
S_A	= the attenuation function (see paper by G. Crisson referred to above).
S_F	= the distribution function (see paper by G. Crisson referred to above).
S	$= S_H + S_W - S_A + S_F$.
S_1	$= S - S_F = S_H + S_W - S_A = -10 \log_{10} 2I_1^2$.
S_2	= Same as S_1 but for terminal return loss.
S_4	$= S_{11} - 2.5$.
S_5	= end path loss with fixed termination.
S_{11} and S_{12}	= values, respectively, toward the West and the East of active or passive single frequency return loss for $S_F = 0$, made up of intermediate paths only.
S_{21} and S_{22}	= same as S_{11} and S_{12} but for end paths.
S_{31} and S_{32}	= same as S_{11} and S_{12} but for combination of intermediate and end paths.
S_I	= return loss which is actually obtained on a particular case.
S_C	= return loss in a particular case with equipment.
S_E	= equipment return loss.
S_P	$= -20 \log a$ = distribution function of two return losses in series, where each such return loss has a distribution curve of the form $S + S_F$.
S_F'	= function corresponding to S_F after equipment has been added.
$S_Q(b)$	= distribution function of singing points or singing margins; i.e., a normal law with an average of zero and a standard deviation of b.
$S_g(d)$	= distribution functions of active return losses with fixed terminations.
t	$= 10^{-(T/20)}$.
T	= taper in db per repeater section.
t_k	$= 10^{-(L_k/40k)}$.
V	$= 1 - P$.
V_m	$= 1 - P_m$.

$\underset{i}{\mathrm{X}}$ = a symbol indicating that the two quantities in decibels are combined as if their currents added directly.

E.g.,

$$S_{11} \underset{i}{\mathrm{X}} S_{21} = S_{31} \text{ means that } 10^{-(S_{11}/20)} + 10^{-(S_{21}/20)} = 10^{-(S_{31}/20)}.$$

$\underset{p}{\mathrm{X}}$ = a symbol indicating that the two quantities in decibels so connected are combined as if their powers added directly.

E.g.,

$$S_{11} \underset{p}{\mathrm{X}} S_{21} = S_{31} \text{ means that } 10^{-(S_{11}/10)} + 10^{-(S_{21}/10)} = 10^{-(S_{31}/10)}.$$

x, y = variables as used.

θ_{jk} = phase angle of current returned from the j th loading section in the k th repeater section.

Operation of Ultra-High-Frequency Vacuum Tubes

By F. B. LLEWELLYN

Previous electronics analyses are extended by the introduction of more general boundary conditions. The results are applied to the calculation of the rectifying properties of diodes at very high frequencies and to the amplifying properties of negative grid triodes at both low and high frequencies. The effect of space charge on the various capacitances in triodes is discussed, and formulas for the amplification factor and plate impedance are presented in terms of the tube geometry. Finally, a discussion of the input impedance of negative grid triodes is given together with a comparison of the theoretical value with the results of measurements made by several well-known experimenters.

IN the study of the functioning of vacuum tubes at ultra-high frequencies it has been necessary to retrace the steps followed in the early history of vacuum tube performance but with the difference that a microscope for viewing the path at close range must be substituted for the telescope with which the original trail was mapped from afar off. As a result of this microscopic survey, formulas have been developed which are applicable to frequencies so high that the time of flight of the electron across the tube may occupy several cycles of the high-frequency oscillation. In addition to this result, several by-products of the study are found to have a useful application in the low-frequency field and to throw additional light on the multitudinous activities of the electrons and their effect upon the external circuit.

In this respect it is particularly interesting to see the way in which the geometry of the vacuum tube enters into the determination of the amplification factor of negative grid triodes and to compare the results now obtained with the earlier results of such workers as Abraham, King, Schottky, Lane and Van der Bijl. The effect of the negative grid on the transit time of the electrons also yields low-frequency relations in which certain new facts concerning the plate resistance are brought out.

Various papers [1, 2, 3, 4, 5] published within the last few years have dealt with the general problem of vacuum tubes in which the electron transit time is of importance and have derived results which are useful in several practical applications. However, in none of these have the initial relations been general enough to allow more than a very rough application to be made to the most widely used tube of all—the negative grid triode. A paper [2] by the present writer contains certain general conclusions concerning the negative grid tube at very high frequencies and derives expressions for the complex values taken by the amplification factor and plate impedance. While these expressions represent the general trend of the variations, they are based on certain a priori assumptions, as was emphasized in the paper, and hence partake of the telescopic viewpoint of the low-frequency vacuum tube analysis rather than the microscopic viewpoint which is now necessary.

It was with the aim of overcoming this limitation that the work described in the present paper was undertaken. Its successful outcome was made possible by a very simple generalization of the methods described in the references, but one which has such far-reaching consequences that it appears worthwhile to start the analysis at the very beginning, and abbreviate only to the extent that intermediate algebraic steps are omitted because of their unwieldy length.

For convenience, the paper is divided into four parts. In Part I the mathematical analysis is outlined in its fundamental form and general working formulas are developed. In a way similar to low-frequency analysis these formulas may be divided into constant or d.-c. relations, first-order relations, second-order relations, and so on; where the first-order relations apply to a.-c. effects for small amplitudes only, the second-order relations contain rectification and distortion terms; and so on, in exact correspondence with the well known low-frequency relations. Part II contains the solution of the first-order relations expressed in appropriate form for later computations, while Part III does the same thing for second-order relations. Also in Part III, by way of illustration, the effect of frequency on the rectifying properties of parallel plane diodes is discussed. Part IV applies the general first-order and d.-c. solutions to the negative grid triode and shows how its various properties depend on frequency. A discussion of the important effect of active grid loss is included.

In certain cases the same formulas are expressed in several different ways. This is done because of the difficulty of determining the most useful method of expression before a large number of applications shall have been made. In most cases the general equations have been arranged to conform as far as possible with the most widely used modes of expression of the corresponding low-frequency equations. Where a choice of modes of expression is available, both modes have usually been given, it being left to future experience to determine the more useful one. While this procedure results in some repetition, the two modes of expression of the same equation are found in many instances to have their individual advantages, the one being more suitable for one type of application while the other is more particularly adapted to a different application.

Part I—General Analysis

In line with most previous electronic papers consideration is here directed to the behavior of electrons between two parallel planes of practically infinite extent. Differing, however, from earlier works, neither plane is to be regarded as constituting a thermionic emitter or cathode in the general sense. For certain special applications conditions may be chosen so that one of the planes coincides with, and assumes the properties of a zero potential cathode, but in developing the general relations this idea is strictly avoided. It will therefore be premised as a starting point that the velocity and the acceleration of the electrons at one of these two planes are given as initial conditions. It will be found that this generalization completely avoids certain ambiguities which were discussed in the February 1935 issue of the *Proceedings of the Institute of Radio Engineers.*[4] It also allows application of the results to be made to a much wider range of devices, including a fairly rigorous treatment of the negative grid triode.

In the following analysis, and differing from previous references, a change in the units employed has been made so that all quantities are expressed immediately in the practical system of engineering units (amperes, volts, ohms, coulombs, etc.) instead of in the electrostatic and electromagnetic systems. This change has been found to be of great advantage in the use of the equations, since it obviates all necessity for the continuous and irritating transformation of units that accompanies the electrostatic and electromagnetic systems.

The analysis starts with the expression for the total current density

$$I = \rho u + \epsilon \frac{\partial E}{\partial t}, \tag{1}$$

where I is current density, amperes per cm.^{2},
ρ is charge density, coulombs per cm.^{3},
u is electron, or charge velocity, cm. per sec.,
ϵ is permittivity of a vacuum, which is $1/36\pi 10^{11} = 8.85 \times 10^{-14}$ farads/cm.,
E is the electric intensity, volts per cm.

The equation of motion of an electron is

$$eE = kma, \tag{2}$$

where e is electronic charge, coulombs,
m is electronic mass, grams,

$e/m = -1.77 \times 10^8$ coulombs per gm.,*
$k = 10^{-7}$ is the ratio of dyne-cm. to joules,
a is electron acceleration, cm. per sec.²

In this equation the effect of a magnetic field is disregarded. This is thoroughly justified until electron velocities approach that of light or the spacing between the two parallel planes becomes comparable with the wave-length of any alternating field considered. A more detailed discussion is given by Benham.[1]

A third fundamental equation is

$$\text{div}\, \epsilon E = \rho,$$

which, for the parallel planes now considered, becomes

$$\epsilon \frac{\partial E}{\partial x} = \rho. \tag{3}$$

From (1), (2) and (3) is obtained

$$\frac{eI}{km\epsilon} = \frac{da}{dt}. \tag{4}$$

The total current, I, may be considered to be composed of two parts, the first being a constant component and the second being a function of time only. On this basis we can write

$$\frac{eI}{km\epsilon} = K + \varphi'''(t), \tag{5}$$

where K is the constant part, and $\varphi'''(t)$ is the variable part, the primes denoting derivatives with respect to the argument in parentheses. Inserting (5) into (4) and integrating once with respect to time, we find:

$$a = K(t - t_a) + \varphi''(t) - \varphi''(t_a) + a_a + \alpha(t_a), \tag{6}$$

where $a_0 + \alpha(t_a)$ is the acceleration when $t = t_a$ and a_a is independent of t_a.

Another integration gives

$$u = K\frac{(t - t_a)^2}{2} + \varphi'(t) - \varphi'(t_a) - (t - t_a)\varphi''(t_a)$$
$$+ (t - t_a)a_a + (t - t_a)\alpha(t_a) + u_a + \mu(t_a), \tag{7}$$

where $u_a + \mu(t_a)$ is the velocity when $t = t_a$ and u_a is independent of t_a.

* This value is based on deflection measurements (which are applicable to vacuum tube analysis) rather than on spectroscopic measurements which give 1.76 $\times 10^8$. Compton and Langmuir[6] use the spectroscopic figure. For a comprehensive discussion of values of physical constants, see Birge, *Phys. Rev. Supp.*, Vol. 1, July, 1929.

A third integration gives

$$x = K\frac{(t-t_a)^3}{6} + \varphi(t) - \varphi(t_a) - (t-t_a)\varphi'(t_a) - \frac{(t-t_a)^2}{2}\varphi''(t_a)$$
$$+ \frac{(t-t_a)^2}{2}a_a + \frac{(t-t_a)^2}{2}\alpha(t_a) + (t-t_a)u_a + (t-t_a)\mu(t_a), \quad (8)$$

where x is zero when $t = t_a$.

This choice of initial conditions allows one of the two parallel planes to be located at the position where x is zero and where the electron velocities and accelerations are given as above in terms of the time instant t_a when the electron crosses the plane, which may be referred to as the " a " plane. When these initial conditions are specified, then (6), (7), and (8) allow the acceleration, velocity and position, respectively, to be determined at any time t, thereafter.

These quantities are expressed in terms of t and t_a whereas it is more convenient in vacuum tube work to have the acceleration and velocity expressed in terms of t and x. Ideally, this could be done by solving (8) for t_a and thence eliminating t_a from (6) and (7). Practically, (8) cannot be solved directly for t_a because it is a higher order equation and involves φ, α and μ which may be (and usually are) transcendental functions of t_a. However, an indirect method can be employed.

If φ, α and μ were zero, then x would be given by the relatively simple equation

$$x = K\frac{(t-t_a)^3}{6} + a_a\frac{(t-t_a)^2}{2} + u_a(t-t_a). \quad (9)$$

Although this is a cubic, nevertheless $(t - t_a)$ may be obtained with relative ease in any particular case. The use of a new variable T to replace x is suggested by (9) and accordingly the defining equation of T under all conditions is taken to be:

$$x = K\frac{T^3}{6} + a_a\frac{T^2}{2} + u_aT, \quad (10)$$

which holds even when φ, α and μ are not zero. It is evident when φ, α and μ are small that T does not differ very much from $t - t_a$ as (10) must then become nearly equivalent to (9). It thus seems expedient to write in general

$$t - t_a = T + \delta, \quad (11)$$

and note that δ becomes very small when φ, α and μ are small.

On the basis of (11), functions of $(t - t_a)$ may be expanded into series in powers of δ as follows:

$$f(t-t_a) = f(T+\delta) = f(T) + f'(T)\delta + \frac{1}{2!}f''(T)\delta^2 + \cdots, \quad (12)$$

and similarly functions of t_a may be written

$$f(t_a) = f(t - T - \delta) = f(t - T) - f'(t - T)\delta + \frac{1}{2!}f''(t - T)\delta^2 - \cdots. \quad (13)$$

When (10), (11), (12) and (13) are used in conjunction with (8) the result is a relation between t, T and δ as follows:

$$\begin{aligned}
0 = {} & \frac{K}{6}[3T^2\delta+3T\delta^2+\delta^3] \\
& +\frac{a_a}{2}[2T\delta+\delta^2]+u_a\delta \\
& +\varphi(t)-\left[\varphi(t-T)-\varphi'(t-T)\delta+\frac{1}{2!}\varphi''(t-T)\delta^2-\cdots\right] \\
& -(T+\delta)\left[\varphi'(t-T)-\varphi''(t-T)\delta+\frac{1}{2!}\varphi'''(t-T)\delta^2-\cdots\right] \\
& -\frac{1}{2}(T^2+2T\delta+\delta^2)\left[\varphi''(t-T)-\varphi'''(t-T)\delta+\frac{1}{2!}\varphi''''(t-T)\delta^2-\cdots\right] \\
& +\frac{1}{2}(T^2+2T\delta+\delta^2)\left[\alpha(t-T)-\alpha'(t-T)\delta+\frac{1}{2!}\alpha''(t-T)\delta^2-\cdots\right] \\
& +(T+\delta)\left[\mu(t-T)-\mu'(t-T)\delta+\frac{1}{2!}\mu''(t-T)\delta^2-\cdots\right]. \quad (14)
\end{aligned}$$

This equation may be written in the form of a power series in δ. It cannot yet be solved directly for δ. It has the advantage, however, that δ is not involved in the transcendental functions φ, α and μ, so that an indirect method may be used. Let δ, φ, α and μ each be split up into series as follows:

$$\left.\begin{aligned}
\delta &= \delta_1 + \delta_2 + \delta_3 + \text{etc.} \\
\varphi &= \varphi_1 + \varphi_2 + \varphi_3 + \text{etc.} \\
\alpha &= \alpha_1 + \alpha_2 + \alpha_3 + \text{etc.} \\
\mu &= \mu_1 + \mu_2 + \mu_3 + \text{etc.}
\end{aligned}\right\} \quad (15)$$

These are to be substituted into (14) and the resulting expression may then be expressed as an infinite series of separate equations such that the first equation includes all linear terms which have the subscript 1, but no other terms; The second equation includes all linear terms having the subscript 2 and also all quadratic terms having the subscript 1; the third equation includes all linear terms having the subscript 3, cubic terms with subscript 1, and also products of quadratic terms with subscript 1 and linear terms with subscript 2. The rules for succeeding equations are analogous, so that in general, the sum of the subscripts of each term of the nth equation is equal to n.

The first of these equations may be solved for δ_1, the second for δ_2, the third for δ_3, and so on, giving the following:

$$\delta_1 = -\frac{\left[\varphi_1(t) - \varphi_1(t-T) - T\varphi_1'(t-T) - \frac{1}{2}T^2\varphi_1''(t-T) + \frac{1}{2}T^2\alpha_1(t-T) + T\mu_1(t-T)\right]}{K\frac{T^2}{2} + a_aT + u_a}, \quad (16)$$

$$\delta_2 = -\frac{\left[\begin{array}{l}\varphi_2(t) - \varphi_2(t-T) - T\varphi_2'(t-T) - \frac{1}{2}T^2\varphi_2''(t-T) \\ \quad + \frac{1}{2}T^2\alpha_2(t-T) + T\mu_2(t-T) \\ \quad + \delta_1\left(\frac{1}{2}T^2\varphi_1'''(t-T) - \frac{1}{2}T^2\alpha_1'(t-T) \right. \\ \qquad \left. + T\alpha_1(t-T) - T\mu_1'(t-T) + \mu_1(t-T)\right) \\ \quad + \delta_1^2\left(K\frac{T}{2} + \frac{1}{2}a_a\right)\end{array}\right]}{K\frac{T^2}{2} + a_aT + u_a}, \quad (17)$$

$$\delta_3 = -\frac{\left[\begin{array}{l}\varphi_3(t) - \varphi_3(t-T) - T\varphi_3'(t-T) - \frac{1}{2}T^2\varphi_3''(t-T) \\ \quad + \frac{1}{2}T^2\alpha_3(t-T) + T\mu_3(t-T) \\ \quad + \delta_2\left(\frac{1}{2}T^2\varphi_1'''(t-T) - \frac{1}{2}T^2\alpha_1'(t-T) \right. \\ \qquad \left. + T\alpha_1(t-T) - T\mu_1'(t-T) + \mu_1(t-T)\right) \\ \quad + \delta_1\left(\frac{1}{2}T^2\varphi_2'''(t-T) - \frac{1}{2}T^2\alpha_2'(t-T) \right. \\ \qquad \left. + T\alpha_2(t-T) - T\mu_2'(t-T) + \mu_2(t-T)\right) \\ \quad + \delta_1^2\left(-\frac{1}{4}T^2\varphi_1''''(t-T) + \frac{1}{2}T\varphi_1'''(t-T) \right. \\ \quad + \frac{1}{4}T^2\alpha_1''(t-T) - T\alpha_1'(t-T) + \frac{1}{2}\alpha_1(t-T) \\ \qquad \left. + \frac{1}{2}T\mu_1''(t-T) - \mu_1'(t-T)\right) \\ \quad + \delta_1\delta_2\left(KT + a_a\right) + \delta_1^3\frac{K}{6}\end{array}\right]}{K\frac{T^2}{2} + a_aT + u_a}, \quad (18)$$

$$\delta_4 = \cdots. \quad (19)$$

These formulas together with (11) and (15) allow the acceleration (6) and velocity (7) to be written in terms of t and T, and hence effectively in terms of t and x, for T is a function of x only, as given by (10). The resulting equations for acceleration and velocity are conveniently broken up into a series of equations in accord with the same rules formulated for obtaining the δ's from (14). Thus, we write

$$a = a_0 + a_1 + a_2 + a_3 + \text{etc.}, \tag{20}$$

where a_0 comprises those terms having zero or no subscript, and obtain from (6):

$$a_0 = KT + a_a, \tag{21}$$

$$a_1 = K\delta_1 + \varphi_1''(t) - \varphi_1''(t-T) + \alpha_1(t-T), \tag{22}$$

$$a_2 = K\delta_2 + \varphi_2''(t) - \varphi_2''(t-T) + \varphi_1'''(t-T)\delta_1 + \alpha_2(t-T) - \alpha_1'(t-T)\delta_1, \tag{23}$$

$$a_3 = K\delta_3 + \varphi_3''(t) - \varphi_3''(t-T) + \varphi_1'''(t-T)\delta_2 + \varphi_2'''(t-T)\delta_1 - \frac{1}{2}\varphi_1''''(t-T)\delta_1^2 + \alpha_3(t-T) - \alpha_2'(t-T)\delta_1 - \alpha_1'(t-T)\delta_2 + \frac{1}{2}\alpha_1''(t-T)\delta_1^2, \tag{24}$$

$$a_4 = \cdots. \tag{25}$$

Treating the velocity in a similar way we write

$$u = u_0 + u_1 + u_2 + u_3 + \text{etc.} \tag{26}$$

and obtain from (7)

$$u_0 = K\frac{T^2}{2} + a_aT + u_a, \tag{27}$$

$$u_1 = KT\delta_1 + a_a\delta_1 + \varphi_1'(t) - \varphi_1'(t-T) - T\varphi_1''(t-T) + T\alpha_1(t-T) + \mu_1(t-T), \tag{28}$$

$$u_2 = KT\delta_2 + \frac{K}{2}\delta_1^2 + a_a\delta_2 + \varphi_2'(t) - \varphi_2'(t-T) - T\varphi_2''(t-T) + T\varphi_1'''(t-T)\delta_1 + T\alpha_2(t-T) + \alpha_1(t-T)\delta_1 - T\alpha_1'(t-T)\delta_1 + \mu_2(t-T) - \mu_1'(t-T)\delta_1, \tag{29}$$

$$u_3 = KT\delta_3 + K\delta_1\delta_2 + a_a\delta_3 + \varphi_3'(t) - \varphi_3'(t-T) + \frac{1}{2}\varphi_1'''(t-T)\delta_1^2 - T\varphi_3''(t-T) + T\varphi_1'''(t-T)\delta_2$$

$$+ T\varphi_2'''(t - T)\delta_1 - \frac{1}{2} T\varphi_1''''(t - T)\delta_1^2 + T\alpha_3(t - T)$$
$$- T\alpha_1'(t - T)\delta_2 - T\alpha_2'(t - T)\delta_1 + \frac{1}{2} T\alpha_1''(t - T)\delta_1^2$$
$$+ \alpha_1(t - T)\delta_2 + \alpha_2(t - T)\delta_1 - \alpha_1'(t - T)\delta_1^2 + \mu_3(t - T)$$
$$- \mu_1'(t - T)\delta_2 - \mu_2'(t - T)\delta_1 + \frac{1}{2} \mu_1''(t - T)\delta_1^2. \tag{30}$$

Aside from their length these equations are not complicated and in applications to special cases many terms are apt to vanish, leaving relatively compact expressions.

In circuit work the potential difference between the two parallel planes, " a " and " b," say, is more often required than the electron acceleration. This may be found from the definition of the potential difference, namely

$$V_a - V_b = \int_a^b E\partial x, \tag{31}$$

in which t remains constant during the integration. From (10)

$$\partial x = \left(K \frac{T^2}{2} + a_a T + u_a \right) dT = u_0 dT, \tag{32}$$

so that with the aid of this equation and (2) the potential difference is given by:

$$W_a - W_b = \int_a^b a dx = \int_0^T a \left(K \frac{T^2}{2} + a_a T + u_a \right) dT, \tag{33}$$

where the symbol, W, is used as an abbreviation for eV/km.

In the same way as the acceleration and velocity are divided into components, the potential difference may be split up into $(W_a - W_b)_0$, $(W_a - W_b)_1$, $(W_a - W_b)_2$, etc. which are defined by:

$$(W_a - W_b)_0 = \int_0^T a_0 u_0 dT, \tag{34}$$

$$(W_a - W_b)_1 = \int_0^T a_1 u_0 dT, \tag{35}$$

$$(W_a - W_b)_2 = \int_0^T a_2 u_0 dT, \tag{36}$$

$$(W_a - W_b)_3 = \int_0^T a_3 u_0 dT, \tag{37}$$

and where the δ's given by (16), (17), (18), etc., are to be inserted in the a's given by (21), (22), (23), etc., before the integration can be carried out.

The formal solution of the problem has been reached with the attainment of (34), (35), (36), etc. As soon as specific functions are chosen for the current, $[K + \varphi'''(t)]km\epsilon/e$, the initial acceleration, $a_a + \alpha(t)$, and the initial velocity, $u_a + \mu(t)$, the integrations can be performed to give the potential difference between two planes located respectively at $x = 0$ and $x = x$.

While the general relation between current and potential is non-linear, the first order current $\varphi'''(t)e/km\epsilon$ is linearly related to the first order potential difference $(W_a - W_b)e/km$, and the results for this case can be expressed conveniently by using complex functions in the manner usual with electrical engineers. This will be done in some of the following applications.

In the treatment of second and higher-order components, it is convenient to select the current components to correspond with powers of the potential rather than vice versa. For example, $(V_a - V_b)_1$ is taken to be the complete expression for the fluctuating component of potential, thus causing $(V_a - V_b)_2$, $(V_a - V_b)_3$, etc. to vanish. Equation (36) then yields an expression for the second-order current in terms of the first-order current, and (37) does likewise for the third-order current.

The general equations are, however, applicable to any converging method of selecting the components, and the proper one for any particular case is to be determined by considerations of simplicity and convenience.

PART II—FIRST-ORDER SOLUTION

This is the linear case, so that to each component of current, $A \sin \omega t$, say, there corresponds a potential component of the form $B \sin (\omega t + \eta)$. It follows that a current of the form Ae^{pt} will produce a corresponding potential difference Pe^{pt}, and that p may be taken to be a generalized exponent having the value $i\omega$ when sinusoidal currents are considered. In the latter case, P will usually be complex. The generalized exponent p results in a more compact symbolism than would be possible with the imaginary exponent, $i\omega$.

Thus, we write for the current

$$\frac{eI}{km\epsilon} = K + \varphi_1'''(t) = K + Je^{pt}. \tag{38}$$

In a similar way the initial fluctuating components of acceleration and velocity are taken to be

$$\alpha_1(t - T) = Ge^{p(t-T)} = (Ge^{-pT})e^{pt}, \tag{39}$$

$$\mu_1(t - T) = He^{p(t-T)} = (He^{-pT})e^{pt}. \tag{40}$$

With this nomenclature, all terms appearing in the first-order equations will contain the factor e^{pt}, which may accordingly be omitted throughout. Thus, instead of writing Ge^{pt} for the acceleration at the " a " plane, the single symbol α_1 will be used, where the omission of the functional notation $\alpha_1(t - T)$ indicates that the acceleration is taken at the " a " plane where T is zero, and that the multiplying factor e^{pt} is understood. In a similar way the symbol μ_1 will indicate the first-order fluctuating component of velocity at the " a " plane with the factor e^{pt} understood.

A still further symbolism will be of assistance. The quantity pT will be denoted by β. When p is the imaginary $i\omega$ then ωT is the transit angle, θ, which has been defined in previous papers,[1, 2] and in the sinusoidal case $\beta = i\theta$.

With this nomenclature, the first-order potential difference, (35), acceleration, (22), and velocity (28) may be written:

$$\begin{aligned}(W_a - W_b)_1 = \frac{J}{p^4}\Bigg[K\left(\frac{\beta^3}{6} - \beta - 2e^{-\beta} - \beta e^{-\beta} + 2\right) \\ + pa_a\left(\frac{\beta^2}{2} + \beta e^{-\beta} + e^{-\beta} - 1\right) + p^2u_a(\beta + e^{-\beta} - 1)\Bigg] \\ - \frac{\alpha_1}{p^2}\left[a_a(\beta e^{-\beta} + e^{-\beta} - 1) + u_a p(e^{-\beta} - 1)\right] \\ + \frac{\mu_1}{p^2}\left[K(\beta e^{-\beta} + e^{-\beta} - 1)\right],\end{aligned} \tag{41}$$

$$\begin{aligned}a_1 = \frac{J}{p}\left[1 - e^{-\beta} - \frac{K}{p^2u_0}\left(1 - e^{-\beta} - \beta e^{-\beta} - \frac{\beta^2}{2}e^{-\beta}\right)\right] \\ + \alpha_1\left(1 - \frac{KT^2}{2u_0}\right)e^{-\beta} - \mu_1\frac{KT}{u_0}e^{-\beta},\end{aligned} \tag{42}$$

$$\begin{aligned}u_1 = \frac{J}{p^2}\left[1 - e^{-\beta} - \beta e^{-\beta} - \frac{a_0}{pu_0}\left(1 - e^{-\beta} - \beta e^{-\beta} - \frac{\beta^2}{2}e^{-\beta}\right)\right] \\ + \frac{\alpha_1}{p}\left(1 - \frac{a_0T}{2u_0}\right)\beta e^{-\beta} + \mu_1\left(1 - \frac{a_0T}{u_0}\right)e^{-\beta}.\end{aligned} \tag{43}$$

An alternate method of procedure, which is particularly useful at moderately low frequencies where the transit angle θ is only a few radians, is to expand functions of $(t - T)$ into power series in T. The same result may be obtained from (41), (42) and (43) above by writing $e^{-\beta}$ in series form. The result is:

$$(W_a - W_b)_1 = \sum_{n=0}^{\infty} (-\beta)^n \left[J\left(\frac{KT^4(n+2)}{(n+4)!} + \frac{a_a T^3(n+2)}{(n+3)!} + \frac{u_a T^2}{(n+2)!} \right) + \alpha_1 \left(\frac{a_a T^2(n+1)}{(n+2)!} + \frac{u_a T}{(n+1)!} \right) - \mu_1 \left(\frac{KT^2(n+1)}{(n+2)!} \right) \right], \tag{41a}$$

$$a_1 = \sum_{n=0}^{\infty} \frac{(-\beta)^n}{n!} \left[JT\left(\frac{1}{n+1} - \frac{KT^2}{2u_0} \frac{1}{(n+3)} \right) + \alpha_1 \left(1 - \frac{KT^2}{2u_0} \right) - \mu_1 \left(\frac{KT}{u_0} \right) \right], \tag{42a}$$

$$u_1 = \sum_{n=0}^{\infty} \frac{(-\beta)^n}{n!} \left[JT^2\left(\frac{1}{n+2} - \frac{a_0 T}{2u_0} \frac{1}{(n+3)} \right) + \alpha_1 T \left(1 - \frac{a_0 T}{2u_0} \right) + \mu_1 \left(1 - \frac{a_0 T}{u_0} \right) \right]. \tag{43a}$$

These latter forms of expression clearly show what happens at low frequencies and they prove that the equations do not give infinite values for any of the components so long as a_0T/u_0 and KT^2/u_0 remain finite. Now in general

$$\frac{a_0 T}{u_0} = \frac{(KT + a_a)T}{K\frac{T^2}{2} + a_a T + u_a}$$

and

$$\frac{KT^2}{u_0} = \frac{KT^2}{K\frac{T^2}{2} + a_a T + u_a},$$

and these remain finite for all finite positive values of T, K, a_a and u_a. Thus the difficulty at the origin which was discussed [4] in the *Proceedings of the Institute of Radio Engineers*, February, 1935 is overcome by the generalized definition of T in (10). It is true that δ_1 still tends toward infinity when T approaches zero if a_a and u_a are zero and μ_1 is different from zero. This means that the ratio of the variation in transit time, δ_1, to the transit time T tends toward infinity when T approaches zero, and when variations in initial velocity are still present with no constant initial acceleration. This is a logical and expected result, but it leads also to the conclusion that the electrons actually halt their forward motion and reverse their direction. When this happens, (1) is no longer applicable because the velocity at a given point has become multi-valued. This restriction was pointed out by Müller [3] as limiting any analysis which starts with (1).

It may be concluded then that the inherent limitation on (41), (42) and (43) is a singly valued velocity rather than the behavior of the δ's in the neighborhood of a point where the acceleration is zero.

Part III—Second-Order Solution

The second-order solution has practical utility in the computation of distortion and of the modulating and detecting properties of ultra-high-frequency thermionic systems. Even in low-frequency applications such computations are long and tedious. The introduction of the added complication of appreciable transit angles causes further difficulty because of the unwieldy length of the equations. Accordingly, instead of a general exposition of second-order effects, a greatly simplified special case will be treated at the present time, leaving the details of a more general solution until the need for it has become more acute.

The rectifying properties of a parallel plane diode operating with complete space charge will be calculated. The complete space charge condition is defined by placing initial velocities and accelerations equal to zero. When this has been done in (16), (17) and (23) and the resulting values of the δ's have been substituted in (23), functions of $(t - T)$ may be expanded into power series to give the following:

$$a_2 = -2\sum_{n=0}^{\infty}\frac{(-T)^{n+1}\varphi_2^{(n+3)}(t)}{(n+3)(n+1)!} + \frac{1}{K}\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{(-T)^{n+m+1}\varphi_1^{(n+3)}(t)\,\varphi_1^{(m+3)}(t)}{(n+3)(m+3)n!m!}. \tag{44}$$

The second-order potential difference is given by (36) where $u_0 = KT^2/2$ for complete space charge. Thus, from (44)

$$(W_a - W_b)_2 = K\sum_{n=0}^{\infty}\frac{(n+2)}{(n+4)!}(-T)^{n+5}\varphi_2^{(n+3)}(t) - \frac{1}{2}\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}\frac{(-T)^{n+m+4}\varphi_1^{(n+3)}(t)\,\varphi_1^{(m+3)}(t)}{(n+m+4)(n+3)(m+3)n!m!}. \tag{45}$$

The second-order potential difference $(W_a - W_b)_2$ will now be taken as zero, which implies that the first-order potential difference only is impressed on the diode. This gives from (45)

$$K\left[\frac{1}{12}T^4\varphi_2'''(t) - \frac{1}{40}T^5\varphi_2^{\text{iv}}(t) + \frac{1}{180}T^6\varphi_2^{\text{v}}(t) - \cdots\right]$$
$$= \frac{1}{2}\left[\frac{T^4\varphi_1'''(t)\varphi_1'''(t)}{36} + \frac{T^6\varphi_1^{\text{iv}}(t)\varphi_1^{\text{iv}}(t)}{96} + \frac{T^8\varphi_1^{\text{v}}(t)\varphi_1^{\text{v}}(t)}{800} + \cdots\right]$$
$$+ \left[-\frac{T^5\varphi_1'''(t)\varphi_1^{\text{iv}}(t)}{60} + \frac{T^6\varphi_1'''(t)\varphi_1^{\text{v}}(t)}{180} - \frac{T^7\varphi_1'''(t)\varphi_1^{\text{vi}}(t)}{756}\right.$$
$$\left. + \frac{T^8\varphi_1'''(t)\varphi_1^{\text{vii}}(t)}{4032} - \frac{T^7\varphi_1^{\text{iv}}(t)\varphi_1^{\text{v}}(t)}{224} + \frac{T^8\varphi_1^{\text{iv}}(t)\varphi_1^{\text{vi}}(t)}{1152} - \cdots\right]. \quad (46)$$

If the first-order voltage which is impressed on the diode is taken to be a single sine wave, it follows from the linearity of the first-order relations that the first-order current is likewise a single sine wave of the same angular frequency, ω. Thus, let

$$\varphi_1'''(t) = A \sin \omega t. \quad (47)$$

Then

$$\varphi_1^{\text{iv}}(t) = A\omega \cos \omega t,$$
$$\varphi_1^{\text{v}}(t) = -A\omega^2 \sin \omega t,$$
$$\varphi_1^{\text{vi}}(t) = -A\omega^3 \cos \omega t.$$

When these are substituted into (46) it is seen that the right hand side of the equation contains a d.-c. term and a double-frequency term. The left hand side must accordingly contain terms of the same frequencies. Hence the most general form which can be assumed for the second-order current $\varphi_2'''(t)e/km\epsilon$ is:

$$\varphi_2'''(t) = (a_0 + a_1\theta + a_2\theta^2 + \cdots)$$
$$+ (b_0 + b_1\theta + b_2\theta^2 + \cdots) \sin 2\omega t$$
$$+ (c_0 + c_1\theta + c_2\theta^2 + \cdots) \cos 2\omega t, \quad (48)$$

where $\theta = \omega T$ is the transit angle, and the a's have no reference to the symbols previously used for acceleration.

When (47) and (48) are substituted into (46) it will be seen that the coefficients of the d.-c. term, the sin $2\omega t$ term and the cos $2\omega t$ term respectively may be equated on the two sides of the equation. This gives three equations and in each of these, coefficients of corresponding powers of the transit angle may be equated, thus providing the values of all of the coefficients in (48).

Without carrying out this procedure in detail it is possible to find the d.-c. component directly from (46) and (47) by noting that time derivatives of the d.-c. component of $\varphi_2'''(t)$ are zero. Hence the left hand side of (46) reduces to its first term. Substitution of (47) into the right hand side and selection of d.-c. components then gives:

$$\varphi_2'''(t)_{\text{d.-c.}} = \frac{1}{12}\frac{A^2}{K}\left[1 - \frac{\theta^2}{40} + \frac{\theta^4}{2800} - \cdots\right]. \qquad (49)$$

This indicates that the rectified current decreases when the transit angle θ becomes appreciable. In the second part of his first electronics paper,[1] Benham reached the conclusion that the rectified current increases with frequency. To reconcile his result with (49) it is only necessary to note that (49) indicates a decrease in rectified current with frequency provided that the amplitude A of the first-order current remains constant, whereas Benham's result was based on a constant amplitude of the first-order voltage. A direct comparison therefore necessitates the computation of the first-order voltage. From (41a) with zero initial acceleration and velocity we have:

$$(W_a - W_b)_1 = \frac{JKT^4}{12}\left[1 - \frac{3}{10}i\theta - \frac{1}{15}\theta^2 + \frac{1}{84}i\theta^3 + \frac{1}{560}\theta^4 \cdots\right]. \qquad (50)$$

Taking the amplitude of the current factor $\varphi_1'''(t)$ to be A as in (47) and the amplitude of the voltage factor $(W_a - W_b)_1$ to be B, we find from (50) that

$$B^2 = A^2\frac{K^2T^8}{144}\left[\left(1 - \frac{1}{15}\theta^2 + \frac{1}{560}\theta^4 - \cdots\right)^2 + \left(-\frac{3}{10}\theta + \frac{1}{84}\theta^3 - \cdots\right)^2\right]$$

or

$$B^2 = A^2\frac{K^2T^4}{144}\left[1 - \frac{13}{300}\theta^2 + \frac{11}{12600}\theta^4 - \cdots\right]. \qquad (51)$$

The A^2 may thus be eliminated between (51) and (49) giving

$$\varphi_2'''(t)_{\text{d.-c.}} = \frac{12B^2}{K^3T^8}\left[1 + \frac{11}{600}\theta^2 + \frac{39}{140000}\theta^4 + \cdots\right], \qquad (52)$$

which is in agreement with the result obtained by Benham.

The physical explanation of the increase in rectified current which occurs when the transit angle becomes appreciable follows directly when it is realized that as the frequency is made higher, the fundamental component of the alternating current increases when the a.-c. voltage is maintained at constant amplitude. This is because of the capacitance of the diode. It is this increase in fundamental current which causes the increase in rectified current, and not the effect of transit angle directly.

Part IV—Negative Grid Triodes

The general aspect of the negative grid triode is shown in Fig. 1. The upper diagram shows a plan view of the electrode arrangement and

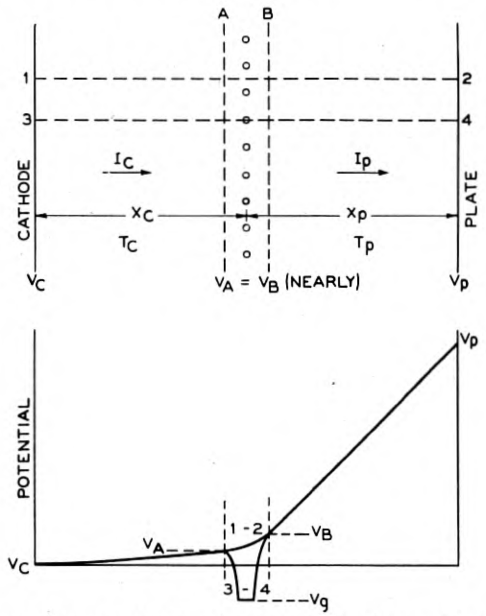


Fig. 1.—Nomenclature and potential distribution in negative grid triodes.

indicates the nomenclature which will be adopted. The planes at V_c and V_p constitute the cathode and plate, respectively, and the grid wires are indicated by the small circles. The planes A and B are imaginary planes located on opposite sides of the grid and only far enough away so that the irregularities in the potential distribution caused by the individual negative grid wires have practically disappeared. An analysis of the potential distribution in the immediate neighborhood of a shielding screen has been made by Maxwell (Treatise on Electricity and Magnetism Vol. I) and this applies to the grid of the triode in the absence of electron flow. The presence of electrons between the negative grid wires will tend to decrease the irregularities shown in Maxwell's analysis and therefore allow the planes A and B to be placed somewhat closer together than his figures would indicate. As far as the writer is aware, exact relations for the potential distribution near the grid wires in the presence of electron flow have not been worked out so that Maxwell's analysis is our only basis at the present time for determining how near to the grid we can imagine A and B to be fixed.

The distance from cathode to grid will be called x_c and from grid to plate will be called x_p. Corresponding to x_c there will be an electron transit time T_c, and to x_p there will be the transit time T_p. The density of the total current leaving the cathode is I_c so that the current factor $eI/km\epsilon$ will be J_c. The factor for the density of the total current reaching the plate will be J_p.

The general scheme for analysis of the triode is to regard the structure as composed of two parallel plane diodes. The first of these comprises the cathode and the imaginary plane A, while the second comprises the imaginary plane B and the plate or anode. In accord with normal operation, complete space charge is postulated at the cathode, so that initial velocities and accelerations are zero, and the general parallel plane equations, (41), (42) and (43) may be applied directly to conditions between the cathode and the plane A.

The plane B and the plate also constitute a structure to which the general equations can be applied, but first the initial velocities and accelerations at B will have to be found, and finally a relation between J_c and J_p will be needed. Moreover, the relation of the potentials at A and B to the potential V_g of the grid wires themselves must be found, because it is only the latter that is available for external use or measurement.

As an aid to finding these various relations the lower diagram in Fig. 1 will be of assistance. This is a graph showing the general form of the d.-c. potential as a function of the distance from the cathode. Between cathode and plane A the potential curve is the same regardless of whether the line 1–2 in the upper diagram which passes between two grid wires is followed, or whether the line 3–4 which passes through a grid wire is followed. The same thing applies between plane B and the plate.

Between planes A and B, however, conditions become vastly different according to whether the potential curve is drawn for the line 1–2 or the line 3–4, and the general shape of the potential curve for the two conditions is marked 1–2 and 3–4 respectively in the lower diagram. Along 1–2 the potential is everywhere positive, as otherwise electrons would not be able to penetrate the grid mesh and reach the plate. On the other hand, the grid wires themselves are at a negative potential, and the curve 3–4 shows the way in which the potential surface forms into pockets surrounding the grid wires. The size of these pockets, and hence the location of the planes A and B is determined by the relative potentials of grid and plate, the size of the grid wires, and the spacing between them.

In applying the general analysis for parallel planes to this type of structure it is possible to find all of the desired initial conditions at the B-plane in terms of conditions at the A-plane as well as to find the actual a.-c. potential of the grid wires themselves and the relation between plate current and cathode current provided that a very simple condition is fulfilled. This condition is merely that the size of the potential pockets surrounding the grid wires is so small that the planes A and B may be taken close enough together to cause the electron transit time between the two to be negligibly small compared with the transit time from cathode to A or from B to the anode. Without this condition, the problem appears almost hopeless. With it, the procedure is straightforward and simple.

For the factors affecting the fulfillment of this condition, Maxwell's analysis fortunately provides a guide that is at least safe, because if the planes can be located near together compared with the distances x_c and x_p on the basis of his analysis, then the smoothing effect of the electrons between grid wires will make the actual operating conditions still better. The adaptation of his equations gives the following expression for the potential distribution and for small diameter grid wires:

$$V = \left[\frac{V_g - \dfrac{x_c}{x_c + x_p} V_p}{\left(\dfrac{x_c x_p}{x_c + x_p} \right) \dfrac{4\pi}{a} - 2 \log_e \left(2 \sin \dfrac{\pi c}{a} \right)} \right] \left[\frac{4\pi x_p}{a} \left(\frac{x_c + d}{x_c + x_p} \right) \right.$$
$$\left. - \log_e \left(1 - 2e^{2\pi d/a} \cos^{2\pi z/a} + e^{4\pi d/a}\right) \right] + V_p \left(\frac{x_c + d}{x_c + x_p} \right). \quad (53)$$

Here d is the distance from the grid to the place where the potential is to be computed, and is considered positive when directed from the grid toward the plate, z is distance measured parallel to the grid, a is the distance between centers of grid wires, c is the wire radius, and the cathode potential, V_c is taken to be zero. The only term here contributing to the potential pockets surrounding grid wires is the logarithmic one which alone involves z, the distance along plane A or B. The condition for the variations to be smoothed out with either positive or negative values of d is

$$e^{2\pi D/a} \gg 2,$$

where D represents the magnitude of d. Below are given some of the values of $e^{2\pi D/a}$

D/a	$e^{2\pi D/a}$
0.1	1.875
0.2	3.514
0.3	6.587
0.4	12.35
0.5	23.17
0.6	43.39
0.7	81.34
0.8	152.5
0.9	285.5
1.0	536.0

This shows that if the magnitude of the cosine term which involves z is to be limited to one per cent of the largest of the three terms constituting the logarithm, then D/a should be greater than 0.84. On the other hand if a ten per cent variation is tolerated, then D/a may be as small as 0.48.

There is a further consideration that tends to smooth out the corrugations in the potential caused by the grid wires. The relation (53) was derived on the assumption that the cathode and plate were quite distant from the grid. When this is not the case, the fact that both are equipotential surfaces tends to crowd the irregularities in toward the grid, and hence to decrease the area of the pockets.

The general conclusion to be drawn from this investigation into the potential distribution is that the electronics analysis can be applied with greatest accuracy to tubes where the grid wires are very close together relative to the distance between grid and either cathode or anode. When this is not the case, the analysis may be expected to show the general trend of the performance in most cases but to be unreliable for quantitative computation. In extreme cases, where the grid wires are very far apart, the grid action approximates more nearly to a change in the effective cathode area than a uniform action over the entire surface. The electronics analysis can be adapted to this extreme case when the change in effective cathode area occurs instantaneously with a change in grid potential, for then the diode analysis applies directly between the effective cathode area and the anode. Means for doing this will become evident when the theory for grids with fine mesh is understood so that the details will not be described.

The analysis for fine-mesh grids now follows, based on the assumption applicable to them that the planes A and B in Fig. 1 are very near together compared with x_c and x_p. As a first consequence of this assumption, the electron velocities at B must be the same as those at A, and hence one of the initial conditions at the B-plane is provided for. A second consequence is that the potential at B is the same as that at A, and therefore the potential between plate and cathode is the sum of the potentials between cathode and A and between B and the plate.

The accelerations at the two planes are not the same. This can be seen from the lower diagram in Fig. 1 when it is remembered that the acceleration is proportional to the slope of the potential curve. The accelerations can be found, however, by a relatively simple calculation and this will be done in the course of the following analysis.

D.-C. Relations

As a preliminary to the treatment of first order effects in negative grid triodes, certain d.-c. relations will be determined.

The distance is related to the transit time by (10). At the cathode and for complete space charge the acceleration and velocity are zero, so that the cathode-grid transit time T_c is given by

$$x_c = K \frac{T_c^3}{6}. \qquad (54)$$

At B the acceleration is yet to be found, but may be taken to be g times that at A. Hence, from (21)

$$a_B = gKT_c. \qquad (55)$$

The velocity at B is the same as that at A, so that from (27)

$$u_B = K \frac{T_c^2}{2}. \qquad (56)$$

These values, (55) and (56), may now be inserted as initial conditions in (10) to give a relation between x_p and T_p. Calling the ratio T_p/T_c by the symbol h and x_p/x_c by the symbol y we thus obtain:

$$gh^2 = y/3 - h - h^3/3. \qquad (57)$$

In this way the acceleration is related to the transit times. These have now to be expressed in terms of quantities that can be measured directly, namely the current and the plate voltage. To do this, the general expression for potential difference is obtained by integration of (34) which gives:

$$(W_a - W_b)_0 = \frac{1}{8}K^2T^4 + \frac{1}{2}Ka_aT^3 + \frac{1}{2}Ku_aT^2 + \frac{1}{2}a_a^2T^2 + a_au_aT. \quad (58)$$

Putting in the initial conditions we have

$$(W_c - W_A)_0 = \frac{1}{8}K^2T_c^4 \quad (59)$$

and

$$(W_B - W_p)_0 = \frac{1}{8}K^2T_c^4[4gh(1 + h^2) + 4g^2h^2 + 2h^2 + h^4]. \quad (60)$$

From these, and placing the cathode potential equal to zero, we get

$$\sqrt{\frac{V_p}{V_A}} = 2gh + 1 + h^2.$$

With (57) the ratio g may be eliminated from this, giving

$$\sqrt{\frac{V_p}{V_A}} = \frac{2y}{3h} - 1 + \frac{1}{3}h^2. \quad (61)$$

When h is small, as it normally is, the last term may be disregarded, giving

$$h = \frac{2/3\ y}{1 + \sqrt{V_p/V_A}} \text{ (approximately)}. \quad (61a)$$

Equations (57) and (61) allow g and h to be found in terms of V_p, which can be measured directly, and V_A which can be found in terms of the current and the distance x_c by combining (54) and (59) to give:

$$(W_c - W_A)_0 = \frac{1}{8}K^{2/3}(6x_c)^{4/3}. \quad (62)$$

This is the well-known Child's equation, and when W and K are expressed in terms of voltage and current appears in the usual form

$$I_0 = -\ 2.34 \times 10^{-6}V_A^{3/2}/x_c^2 \text{ amperes/cm.}^2 \quad (63)^*$$

Several other expressions which are useful for computation purposes may be found from these d.-c. relations:

* The negative sign occurs because of the assumed current direction, from the cathode. The numerical factor is 2.34 instead of 2.33 given by Compton and Langmuir [6] because of the value of e/m which was used, q.v.

The transit angle $\theta_c = \omega T_c$ may be written in terms either of the voltage V_A or the current I_0, thus

$$\theta_c = \frac{9500x_c}{\lambda\sqrt{V_A}} = \frac{-126}{\lambda}\left(\frac{x_c}{I_0}\right)^{1/3} \text{radians,} \tag{64}$$

where λ is the free-space wave-length in centimeters of an alternating current of angular frequency ω.

The slope of the static characteristic of a diode coinciding with the cathode and the plane A may be expressed

$$-\left.\frac{\partial V_A}{\partial I_0}\right|_{x_c} = r_c = -\frac{2}{3}\frac{V_A}{I_0} = \frac{285{,}000x_c^2}{\sqrt{V_A}} = \frac{-3780x_c^{4/3}}{I_0^{1/3}}, \tag{65}$$

where r_c is the low-frequency resistance in ohms of a square centimeter of area.

From (64) and (65) it can be seen that

$$r_c = 30\lambda x_c\theta_c. \tag{66}$$

A further expression that occurs frequently in following equations is:

$$\frac{KT_c^4}{\epsilon} = 12r_c. \tag{67}$$

Later we shall be able to show how the low-frequency plate resistance of the triode is related to r_c and the amplification factor, as well as how the inter-electrode capacitances of the " cold " tube are involved in these quantities. A simple approximation for the transit time ratio h will also be derived.

First-Order Relations

In the picture shown by Fig. 1 an alternating current is assumed to flow from the cathode to the plane A. This current I_c is related to the quantity J in the general equations (41) or (41a) by the expression $J_c = eI_c/km\epsilon$, and the current includes both conduction and displacement components. Complete space charge at the cathode allows initial velocities and accelerations to be placed equal to zero, so that the first order potential difference between cathode and plane A is given by (41a) as follows:

$$(W_c - W_A)_1 = J_cKT_c^4 \sum_{n=0}^{\infty} (-\beta_c)^n \frac{(n+2)}{(n+4)!}\cdot \tag{68}$$

In a similar way, the potential between plane B and the plate can be written when the initial first-order acceleration at B has been found, the d.-c. initial conditions being given in the previous section, and the first order initial velocity being the same as that at A. From (43a) we have

$$\mu_B = \mu_A = J_c T_c^2 \sum_{m=0}^{\infty} \frac{(-\beta_c)^m (m+1)}{(m+3)!} . \tag{69}$$

The computation of the a.-c. component of acceleration at B is based on the particular property possessed by negative grid tubes that no electrons reach the negative grid. Then, because the velocity at B is the same as that at A it follows that the conduction component of the current is the same at both planes. The total first-order current may be written from (1)

$$I_1 = (\rho u)_1 + \epsilon \frac{\partial E_1}{\partial t} .$$

Multiplying through by $e/km\epsilon$ we can write

$$J = Q + \frac{\partial a_1}{dt} , \tag{70}$$

where $Q = (\rho u)_1 e/km\epsilon$ measures the conduction current, and is the same at planes A and B. Hence at plane A, and since $\partial/\partial t = p$ for exponential currents and voltages:

$$J_c = Q + p a_{1A} . \tag{71}$$

At plane B we have similarly

$$J_p = Q + p \alpha_B . \tag{72}$$

From (71) and (72) there results

$$p \alpha_B = J_p - J_c + p a_{1A} . \tag{73}$$

The value of a_{1A} may be obtained from (42a) so that the acceleration at B is given by

$$p \alpha_B = J_p - J_c - 2J_c \sum_{m=0}^{\infty} \frac{(-\beta_c)^{m+1}(m+2)}{(m+3)!} . \tag{74}$$

All of the initial values at the plane B have been obtained and are expressed in (55), (56), (69) and (74). The potential between B and the plate is now obtained from (41a) and follows, where $h = T_p/T_c$:

$$\frac{(W_B - W_p)_1}{KT_c^4} = (J_p - J_c)\frac{h(gh+1)}{2\beta_c} + J_p h^4 \sum_{n=0}^{\infty} \frac{(-h\beta_c)^n (n+2)}{(n+4)!}$$

$$+ J_c\left[h^2 \sum_{n=0}^{\infty} (-h\beta_c)^n \left(\frac{2gh(n+2)+(n+3)}{2(n+3)!}\right)\right.$$
$$+ h\sum_{n=0}^{\infty}\sum_{m=0}^{\infty} h^n(-\beta_c)^{n+m} \tag{75}$$
$$\left.\times\left(\frac{2gh(n+1)(m+2)+(n+2)(m+2)-h(n+1)(m+1)}{(n+2)!(m+3)!}\right)\right].$$

The attainment of (68) and (75) does not completely solve the problem because it is the potential of the grid wires that can be measured rather than the potential at A and B. The transformation may be readily accomplished, however, by writing

$$(V_A - V_g)_1 = I_g Z_g, \tag{76}$$

where Z_g is the effective impedance between the plane A (or B) and the grid wires. In the negative grid tube when A and B are close together, Z_g is a pure capacitance, C_g.

Writing (68) and (75) respectively in terms of V and I instead of W and J we now have in symbolic form

$$(V_c - V_A)_1 = I_c Z_c, \tag{77}$$
$$(V_A - V_p)_1 = I_c(Z_3 - Z_1) + I_p(Z_1 + Z_2), \tag{78}$$

where

$$I_c = I_g + I_p. \tag{79}$$

The four equations, (76), (77), (78) and (79) are the basis of negative grid triode analysis. The impedances involved refer to a square centimeter of area and are as follows:

$$Z_g = \frac{1}{pC_g} = \text{impedance between plane } A \text{ (or } B\text{) and the grid wires,} \tag{80}$$

$$Z_c = 12r_c \sum_{n=0}^{\infty} (-\beta_c)^n (n+2)/(n+4)!, \tag{81}$$

$$Z_1 = 12r_c h(gh+1)/2\beta_c, \tag{82}$$

$$Z_2 = 12r_c h^4 \sum_{n=0}^{\infty} (-h\beta_c)^n (n+2)/(n+4)!, \tag{83}$$

$$Z_3 = 12r_c h^2\left[\sum_{n=0}^{\infty}(-h\beta_c)^n\left(\frac{2gh(n+2)+(n+3)}{2(n+3)!}\right)\right]$$
$$+ 12r_c h\left[\sum_{n=0}^{\infty}\sum_{m=0}^{\infty}(-\beta_c)^{n+m}h^n\right.$$
$$\left.\times\left(\frac{2gh(n+1)(m+2)+(n+2)(m+2)-h(n+1)(m+1)}{(n+2)!(m+3)!}\right)\right]. \tag{84}$$

The value of r_c is expressed by (65).

For large values of transit angle it is convenient to have these impedances expressed in closed instead of in series form. By using (41) instead of (41a) throughout the analysis the result may be obtained. The expressions for Z_g and Z_1 will not differ from those given above, but Z_c, Z_2 and Z_3 may be written in the following forms:

$$Z_c = \frac{12r_c}{\beta_c^4}\left[\frac{\beta_c^3}{6} - \beta_c(e^{-\beta_c}+1) - 2(e^{-\beta_c}-1)\right], \tag{81a}$$

$$Z_2 = \frac{12r_c}{\beta_c^4}\left[\frac{h^3\beta_c^3}{6} - h\beta_c(e^{-h\beta_c}+1) - 2(e^{-h\beta_c}-1)\right], \tag{83a}$$

$$\begin{aligned} Z_3 = \frac{12r_c}{\beta_c^4}\Bigg[& \beta_c^3\frac{1}{2}(h+gh^2) \\ & + \beta_c[(h-2gh-1)e^{-(h+1)\beta_c} + he^{-h\beta_c} + e^{-\beta_c}] \\ & + [2(h-gh-g)e^{-(h+1)\beta_c} + 2(1-h+gh)e^{-h\beta_c} + 2ge^{-\beta_c} - 2] \\ & + \frac{2}{\beta_c}(1-g)(e^{-(h+1)\beta_c} - e^{-h\beta_c} - e^{-\beta_c} + 1)\Bigg]. \end{aligned} \tag{84a}$$

Of the various impedances Z_3 is the only one that is really troublesome. The values of Z_c and Z_2 may be obtained from data given in published papers [1, 2] when it is noticed that Z_2/h^4 may be calculated from Z_c if β_c is replaced by $\beta_p = h\beta_c$. In the treatment of Z_3 there seems to be no easy road, although the series form (84) may be expanded with comparative ease.

Further steps consist in the transposition of these equations to obtain convenient forms and to show how they harmonize with low-frequency theory. A useful expression may be obtained from the fundamental relations (76), (77), (78), (79) by eliminating V_A, I_c and I_g. The result is

$$\begin{aligned} (V_c - V_p)_1 + \frac{(Z_1 - Z_c - Z_3)}{Z_c + Z_g}(V_c - V_g)_1 \\ = I_p\left[\frac{[(Z_c + Z_2 + Z_3)Z_g + (Z_1 + Z_2)Z_c]}{Z_c + Z_g}\right]. \end{aligned} \tag{85}$$

This begins to look like the familiar low-frequency equation

$$(V_c - V_p)_1 + \mu(V_c - V_g)_1 = I_pZ_p, \tag{86}$$

where μ is the amplification factor and Z_p is the internal plate impedance. The two are equivalent at all frequencies if I_p is interpreted as the density of the total plate current, and not the conduction component, only. Then we have:

$$\mu = \frac{Z_1 - Z_c - Z_3}{Z_c + Z_g} = \frac{(Z_1 + Z_2) - (Z_c + Z_2 + Z_3)}{Z_c + Z_g}, \tag{87}$$

$$Z_p = \frac{(Z_c + Z_2 + Z_3)Z_g + (Z_1 + Z_2)Z_c}{Z_c + Z_g}. \tag{88}$$

The relations involved in these two equations may be made somewhat clearer by finding what they resolve into at low frequencies. This may be done by going to (80)–(84). The first thing to notice is that Z_1 and Z_g become very large because the frequency term $p = i\omega$ appears in their denominators. The other terms are relatively small at low frequencies and we have:

$$\mu_0 \to Z_1/Z_g, \tag{87a}$$

$$r_p = Z_p \to Z_c(1 + \mu_0) + Z_2 + Z_3. \tag{88a}$$

It will be shown below that this formulation for the amplification factor is in accord with that derived by Maxwell in his "Treatise on Electricity and Magnetism" for the shielding effect of a grid mesh.

Low-Frequency Relations in Negative Grid Triodes

In (87a) and (88a) the general form of the low-frequency triode relations is given. It is instructive to compute these in some detail so that the role played by the capacitance C_g between the planes A or B and the grid wires is demonstrated.

To do this, (82) which gives the impedance Z_1 may first be transformed by aid of (57) and (67) to give

$$Z_1 = \frac{x_p}{\epsilon p}(1 - h^3/y). \tag{89}$$

The impedance Z_g may be written as in (80) so that the low-frequency amplification factor is

$$\mu_0 = \frac{x_p}{\epsilon} C_g(1 - h^3/y). \tag{90}$$

It is of interest to note that ϵ/x_p is the capacitance per unit area between the plate and a solid plane at the grid. As expressed by Compton and Langmuir [6] from Maxwell's analysis the low-frequency amplification factor is

$$\mu_0' = \frac{x_p}{\frac{a}{2\pi}\log_e \frac{a}{2\pi c}}, \tag{91}$$

where a is the distance between centers of grid wires and c is the wire radius.

Comparing (90) and (91) we see that both are proportional to x_p, but that (90) contains a correction term, h^3/y which is normally very small as h is usually of the order of one-third to one-tenth. The presence of the correction term may be explained by remembering that (90) was derived for conditions holding when electrons are flowing while (91) applies strictly only to a cold tube in which no free electrons are present. This being the case, it is to be expected that the two equations would be equivalent if the correction term were omitted from (90). The term being small, this is very nearly true in any case and gives

$$C_g = \frac{2\pi\epsilon}{a \log_e \dfrac{a}{2\pi c}} \text{ farads in cm.}^2 \tag{92}$$

The capacitance C_g which was introduced as existing between the electron stream and the grid wires is thus shown to have a value which can be calculated with fair exactness under the conditions when Maxwell's equation (91) holds, namely when the grid wires are small compared with their separation.

Attention is now directed to the low-frequency value of the plate impedance, Z_p. From (88a) together with (81), (83) and (84) it may be shown that

$$r_p = r_c\left[\mu_0 + \frac{4}{3}(1 + y)(1 + h) - \frac{1}{3}(1 + h)^4\right]. \tag{93}$$

This is thought to be the first instance of an expression for the plate resistance derived on strictly theoretical grounds. The formula as it stands contains r_c which is given by (65) and h which is given by (61), and both involve V_A. This latter may be found from (63) in terms of the direct current. A convenient approximation for V_A is obtained by making the assumption that the presence of electrons changes the d.-c. potential V_A by a small amount only, so that its value may be calculated from the static capacitances of a cold tube. The appropriate diagram is shown on Fig. 2.

From the figure, putting $V_c = 0$ we obtain

$$V_{A0} = \frac{V_{p0} + \dfrac{C_g}{C_p} V_{g0}}{1 + \dfrac{C_g}{C_p} + \dfrac{C_c}{C_p}}.$$

It was shown in (90) that C_g/C_p is the static, or "cut-off" amplification factor, μ_0' of the tube. Again, $y = C_c/C_p$ so that

$$V_{A0} = \frac{V_{p0} + \mu_0' V_{g0}}{1 + \mu_0' + y} \text{ (approximately).}$$

When the cathode is heated so that electrons flow, the potential V_A becomes depressed below this value which should be used only in forming rough estimates.

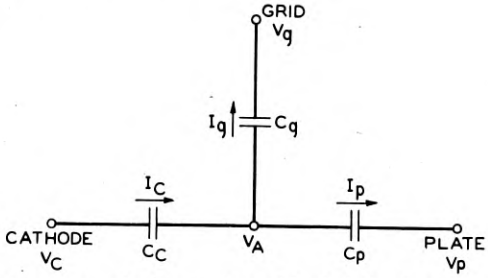


Fig. 2.—Equivalent network of negative grid triode in the absence of electrons.

A more accurate formula for V_A involves the transit time ratio h, but may be obtained as follows: The plate current of an ideal negative grid triode is proportional to $(V_{p0} + \mu V_{g0})^{3/2}$ so that

$$r_p = \frac{2}{3} \frac{(V_{p0} + \mu_0 V_{g0})}{I_0}.$$

Similarly

$$r_c = \frac{2}{3} \frac{V_{A0}}{I_0}.$$

From these two equations together with (93) is obtained

$$V_{A0} = \frac{V_{p0} + \mu_0 V_{g0}}{\mu_0 + \frac{4}{3}(1 + y)(1 + h) - \frac{1}{3}(1 + h)^4},$$

which may also be written

$$V_{A0} = \frac{V_{p0} + \mu_0 V_{g0}}{1 + \mu_0 + \frac{4}{3} y(1 + h) - \frac{1}{3} h^2(6 + 4h + h^2)}. \tag{94}$$

This is in a form for comparison with the approximate equation above and shows the modification produced by the presence of electrons.

General Relations in Negative Grid Triodes

The high-frequency values of the amplification factor and the plate impedance could be computed in detail from (87) and (88) together with the various expressions given for the impedances involved. To do this would require an enormous amount of computation, so that the details are deferred until such time as it becomes evident that (87) and (88) express the high-frequency properties of tubes in the most useful way. The fact that they are analogous to the ordinary low-frequency conventions does not at all assure their general utility in the high-frequency field, and the fundamental equations (76) to (79) may be arranged in a wide variety of forms. For example, a companion equation to (85) may be obtained from the fundamental equations by eliminating V_A, I_p and I_c and thus obtaining:

$$(V_c - V_g)_1 - \left(\frac{Z_c}{Z_2 + Z_3 + Z_c}\right)(V_c - V_p)_1 = I_g\left[\frac{(Z_2 + Z_3 + Z_c)Z_g + (Z_1 + Z_2)Z_c}{Z_2 + Z_3 + Z_c}\right]. \quad (95)$$

Just as (85) gives an equivalent circuit between cathode and plate involving the whole current reaching the plate, so does (95) give an equivalent circuit between cathode and grid involving the whole current reaching the grid. The two equations completely describe the tube performance when the external connections are known. Because of the way in which I_g and I_p are defined they include the so-called grid-plate path, which is treated as a separate circuit at low frequencies.

The impedance presented by a tube to an e.m.f. applied between cathode and grid may now be calculated. To carry out the computation in general, it would be necessary to know the impedance attached between plate and cathode in the external circuit so that $(V_c - V_p)_1$ could be obtained from (85). However, the high-frequency properties of negative grid tubes may be illustrated more directly by choosing for consideration a special case that avoids having to take this additional step involving (85). This special case is the one where such a large capacitance is connected between the plate and cathode that $(V_c - V_p)_1$ is zero for any of the frequencies to be considered. The result will therefore be particularly applicable to finding the input impedance of screen tubes where the requirement for the special case is fulfilled.

For this special case where $(V_c - V_p)_1$ is zero, (95) may be solved directly for the impedance Z_a presented to an input applied between grid and cathode and gives

$$Z_a = Z_g + \frac{(Z_1 + Z_2)Z_c}{Z_2 + Z_3 + Z_c}. \tag{96}$$

The impedance Z_g is a pure capacitance, but the second term on the right of (96) contains both reactive and resistive components which latter account for the active grid loss which has been the subject of several investigations both of a theoretical and experimental nature.[2, 7, 8] At very low frequencies the capacitance represented by Z_1 predominates and the resistive component vanishes leaving for the input impedance Z_a merely the following

$$Z_a \rightarrow Z_g + Z_1\left(\frac{Z_c}{Z_2 + Z_3 + Z_c}\right)$$
$$= \frac{(y - h^3)}{i\omega\mu_0 C_c}\left[\frac{r_p/r_c}{\frac{4}{3}(1 + y)(1 + h) - \frac{1}{3}(1 + h)^4}\right]. \tag{97}$$

This expression may be written in several different forms but in none of them does a simplification occur in the way in which the transit time ratio h enters the equation. Perhaps the best mode of expression is a comparison of the " hot " capacitance C_a given by (97) where $Z_a = 1/i\omega C_a$ with the capacitance C_0 of a cold tube with plate and cathode tied together. This latter may be written

$$C_0 = \frac{\mu_0' C_p(1 + y)}{1 + y + \mu_0'}, \tag{98}$$

where μ_0' is the " cold " amplification factor, (91), and is related to μ_0 as shown by (90) so that $\mu_0 = \mu_0'(1 - h^3/y)$. The ratio C_a/C_0 is the " dielectric constant " of the hot tube and is

$$\frac{C_a}{C_0} = \left(\frac{1 + y + \mu_0'}{1 + y}\right)\left[\frac{\frac{4}{3}(1 + y)(1 + h) - \frac{1}{3}(1 + h)^4}{\mu_0 + \frac{4}{3}(1 + y)(1 + h) - \frac{1}{3}(1 + h^4)}\right]. \tag{99}$$

For illustration suppose that a certain tube has the following values: $y = 1$, $\mu_0 = 10$, $h = 1/5$. Then from (99) the dielectric constant of the hot tube would be 1.19. This is somewhat less than the value of

4/3 obtained by Benham [1] in an analysis which considered effects between cathode and grid only. Benham's value applies strictly to the capacitance between cathode and the plane A in Fig. 1 while (99) applies to the parallel combination of cathode-grid and grid-plate capacitances. The constants of the individual capacitances may be calculated from the fundamental equations (76)–(79) but are omitted here because of space limitations. A series of experiments performed several years ago by Mr. A. J. Rack and the writer and covering a wide range of operating conditions with several vacuum tubes showed values greater than unity for the dielectric constant of the cathode-grid capacitance and less than unity for the grid-plate capacitance. The parallel combination had a dielectric constant which was greater than unity in the range investigated, being thus in accord with (99) which always gives constants greater than unity for normal values of h.

The input capacitance of detector- and voltmeter-tubes being thus a somewhat complicated function, it is to be expected that the calculation of the impedance at higher frequencies where transit times are appreciable will be similarly complicated. To avoid undue length only the first term contributing to the resistive component of the active grid loss will be computed in detail. It must be pointed out, however, that the series in powers of transit angle which represents the input impedance converges slowly so that the first term is useful only when the transit time is small.

Keeping this in mind, we go to (96) and write the impedance in series form, obtaining finally the following expression for the equivalent shunting resistance between cathode and grid:

$$\frac{1}{R_a} = \frac{\theta_c^2}{180}\frac{\mu_0}{r_p}\left[\frac{y^2A - yB + C}{(y - h^3)^2}\right] \times \left[\frac{\mu_0}{\mu_0 + \frac{4}{3}(1+y)(1+h) - \frac{1}{3}(1+h)^4}\right], \tag{100}$$

where

$$A = 9 + 44h + 45h^2,$$
$$B = 51h^2 + 123h^3 + 55h^4 + 3h^5,$$
$$C = 45h^4 + 51h^5 + 24h^6 + 11h^7 + 3h^8.$$

As in previous cases, this may be written in several ways, depending on the mode of expression of θ_c and r_p. For example

$$\theta_c = 2r_c\omega C_c,$$

so that

$$\frac{1}{R_a} = \frac{(\omega C_c)^2}{45} \frac{r_p}{\mu_0} \left[\frac{y^2 A - yB + C}{(y - h^3)^2} \right] \times \left[\frac{\mu_0}{\mu_0 + \frac{4}{3}(1 + y)(1 + h) - \frac{1}{3}(1 + h)^4} \right]^3. \quad (100a)$$

Comparison of (100) and (100a) shows that the transconductance μ_0/r_p appears in the numerator of the former, but in the denominator of the latter, and illustrates the care that must be taken in deriving sweeping conclusions concerning the effect of various tube parameters without taking all of the contributing factors into consideration. In the case of (100) the conclusion is that the loss may be reduced by decreasing the transconductance, but only if the transit angle θ_c is unchanged. On the other hand, (100a) says that the loss may be reduced by increasing the transconductance, but only if this is accomplished without change in the cathode-grid spacing, and without altering h by an amount large enough to affect materially the factors in square brackets.

Experimentally, it is found in many tubes that the loss increases when the transconductance is increased by changing the voltages applied to a given tube. This would seem to be at variance with (100a), for the cathode-grid spacing, and hence C_c, has not been altered by the voltage change. The explanation of the difficulty apparently lies in the departure of the static characteristics of many tubes from the 3/2 power law, which again may be explained in part by the presence of initial velocities and the large size of the potential pockets surrounding the grid wires. In a rough way the action of the latter is to vary the effective cathode area when the voltages are changed, producing an increased area with increase of current, and hence producing a current variation greater than the 3/2 power law.

In a recent paper, D. O. North [11] derives a formula for the active grid loss by neglecting space charge between grid and plate. His result is similar in many respects to (100) and both contain the factors $\theta_c^2\mu_0/r_p$. In an experimental check, W. R. Ferris [10] secures excellent results by obtaining the transconductance from the static characteristics of the tubes used, but computing θ_c by a formula which differs only slightly from (94). It can be shown that such a procedure would give a computed loss which increases with transconductance when the static characteristic is of the form $I = K(V_{p0} + \mu_0 V_{g0})^n$ and when n is greater than 2. The static characteristics are not given in Mr.

Ferris' paper so that it is not evident whether the exponent is greater than 2. The loss did, however, increase with transconductance and was checked by the computations in a satisfactory manner, which would imply either that the exponents were actually greater than 2 in Mr. Ferris' tubes, or that their cylindrical shape caused a decrease in the effect of the transconductance on the transit angle.

The equations in general indicate that the shunting resistance between cathode and grid is proportional to the square of the wave-length. This is in accord with the theory and experiments of Thompson and Ferris,[7, 10] and with the experiments of J. G. Chaffee [8] on tubes biased as class *A* amplifiers. However, when the tubes were biased as detectors, Chaffee [8, 9] found that the resistance varied more nearly as the first power of the wave-length. There are several factors which may contribute to this difference. With detector bias near cut-off the transit angles are large so that more terms of the fundamental equations may be needed. In Chaffee's work these were computed to be of the order of two or three radians which was scarcely enough to cause the entire effect observed by him. Another cause is thought to be an actual reversal in the direction of motion of electrons caused by the alternating potential operating in the vicinity of cut-off. Further study both of experimental and theoretical nature is required, however, before the point can be considered to be satisfactorily explained.

Conclusion

In general, the analysis presented in the foregoing pages is capable of serving as a guide to indicate the kind of results to be expected in the operation of vacuum tubes at ultra-high frequencies. In those cases where the physical structure of the tube complies with the conditions laid down for the theoretical treatment, a quantitative agreement can be anticipated. The importance of departures of the physical structure from this ideal can be evaluated in many instances by a careful comparison of the actual with the ideal structure.

Much yet remains to be done in the way of computing and tabulating the various factors involved in the equations and of investigating the effects of such things as initial electron velocities from a hot cathode, large size grid wires, coarse mesh grids, and cylindrical structures.

References

1. W. E. Benham, "Theory of the Internal Action of Thermionic Systems at Moderately High Frequencies." *Phil. Mag.*, p. 641, March 1928. Part II, *Phil. Mag. Supp.*, Vol. 11, p. 457, February 1931.
2. F. B. Llewellyn, "Vacuum Tube Electronics at Ultra-High Frequencies." *Proc. I.R.E.*, Vol. 21, November 1933; also *B.S.T.J.*, Vol. XIII, January 1934.

3. Johannes Müller, "Elektronenschwingungen im Hochvakuum." *Hochfrequenztechnik und Elektroakustik*, Vol. 41, May 1933.
4. F. B. Llewellyn, "Note on Ultra-High-Frequency Electronics." *Proc. I.R.E.*, Vol. 23, February 1935.
5. W. E. Benham, "Electronic Theory and the Magnetron Oscillator." *Proc. Phys. Soc.*, Vol. 47, January 1935.
6. Compton and Langmuir, "Electrical Discharges in Gases," Part II. *Rev. of Mod. Physics*, Vol. 3, April 1931.
7. B. J. Thompson and W. R. Ferris, Oral paper at Rochester I.R.E., 1934.
8. J. G. Chaffee—Forthcoming discussion on (7).
9. J. G. Chaffee, "The Determination of Dielectric Properties at Very High Frequencies." *Proc. I.R.E.*, Vol. 22, August 1934.
10. W. R. Ferris, "Input Resistance of Vacuum Tubes as Ultra-High Frequency Amplifiers." *
11. D. O. North, "Analysis of the Effects of Space Charge on Grid Impedance." *

* To be published shortly. Much of the material was presented orally by B. J. Thompson and W. R. Ferris at various technical meetings including:

U. R. S. I. and *I. R. E.*, Washington, April 27, 1934.
Philadelphia section of *I. R. E.*, January 3, 1935.
Reference (7) above.

Further Extensions of the Theory of Multi-Electrode Vacuum Tube Circuits

By S. A. LEVIN and LISS C. PETERSON

The response of circuits containing vacuum tubes with any number of electrodes due to impressed electromotive forces, and under such circumstances that the time of transit of the electrons is negligible, is discussed when arbitrary feedback is present between the circuits connected to the electrodes, each of which may carry conductive current. The use of the theory is illustrated by obtaining first and second order effects in typical three-electrode tube circuits.

In a previous paper in the *Bell System Technical Journal*, October, 1934, the treatment was restricted to three-electrode tube circuits in which it was assumed that the amplification factor of the tube was constant and that no conductive grid current was present. In the present paper these restrictions are removed.

Introduction

THE response in multi-electrode vacuum tube circuits due to impressed electromotive forces has been the subject of several papers. For the three-electrode vacuum tube circuit J. R. Carson [1] has used a method of successive approximations, assuming constant amplification factor and no conductive grid current. E. Peterson and H. P. Evans [2] removed the restriction on the amplification factor but maintained the assumption regarding the grid current, while F. B. Llewellyn [3] considered the general case with both plate and grid currents. Finally, J. G. Brainerd [4] has treated the general case of the four-electrode tube circuit. The theories given by these authors did not take into account any feedback between the circuits of the electrodes except in the first approximation.

In a previous paper [5] the theory given by Carson was extended to include the effects of feedback between plate and grid circuits not only in the first but also in the second and higher approximations. The aim of the present paper is to extend similarly the other theoretical work mentioned above[2, 3, 4] to circuits containing tubes with three, four, or any number of electrodes.

Theory of Three-Electrode Tube Circuits

We shall consider the three-electrode tube circuit shown in Fig. 1 where Z_1, Z_2, and Z_3 are impedances which may include inter-electrode

[1] J. R. Carson: *I. R. E. Proc.*, April, 1919, p. 187.
[2] E. Peterson and H. P. Evans: *B. S. T. J.*, July, 1927, p. 442.
[3] F. B. Llewellyn: *B. S. T. J.*, July, 1926, p. 433.
[4] J. G. Brainerd: *I. R. E. Proc.*, June, 1929, p. 1006.
[5] S. A. Levin and Liss C. Peterson: *B. S. T. J.*, October, 1934, p. 523.

admittances. The impressed variable electromotive forces are ϵ_p and ϵ_g in series with the impedances Z_p and Z_g, respectively. We will designate by E_p and I_p the total plate voltage and current, respectively, while the corresponding quantities for the grid are E_g and I_g. In the

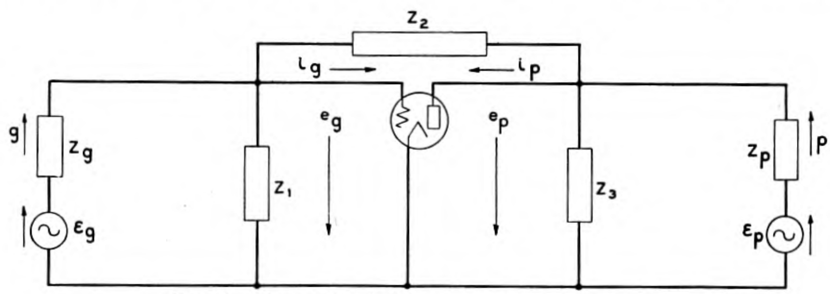


Fig. 1—Three-electrode vacuum tube and circuit.

absence of the variable electromotive forces the d-c. values of these voltages and currents are E_{p0}, I_{p0}, E_{g0}, and I_{g0}, respectively, while the increments due to the impressed forces are e_p, i_p, e_g, and i_g, respectively. Similarly, g and p denote the incremental voltages across Z_g and Z_p. All quantities referring to currents and voltages are instantaneous values.

We will now assume that I_p and I_g are functions of E_p and E_g, and that we can derive from these functions the expansions

$$i_p = \sum b_{mn} e_p{}^m e_g{}^n, \qquad i_g = \sum \beta_{mn} e_p{}^m e_g{}^n, \tag{1}$$

where

$$b_{mn} = \frac{1}{m!n!}\frac{\partial^{(m+n)} I_p}{\partial E_p{}^m \partial E_g{}^n}, \qquad \beta_{mn} = \frac{1}{m!n!}\frac{\partial^{(m+n)} I_g}{\partial E_p{}^m \partial E_g{}^n}, \tag{2}$$

evaluated at the operating point (E_{p0}, E_{g0}).

The important tube parameters are by definition

$$\left.\begin{aligned} \frac{1}{r_p} &= \frac{\partial I_p}{\partial E_p}, \quad \mu_p = \frac{\dfrac{\partial I_p}{\partial E_g}}{\dfrac{\partial I_p}{\partial E_p}} = -\left(\frac{dE_p}{dE_g}\right)_{I_p=\text{const.}} S_{pg} = \frac{\partial I_p}{\partial E_g} = \frac{\mu_p}{r_p} \\ \frac{1}{r_g} &= \frac{\partial I_g}{\partial E_g}, \quad \mu_g = \frac{\dfrac{\partial I_g}{\partial E_p}}{\dfrac{\partial I_g}{\partial E_g}} = -\left(\frac{dE_g}{dE_p}\right)_{I_g=\text{const.}} S_{gp} = \frac{\partial I_g}{\partial E_p} = \frac{\mu_g}{r_g} \end{aligned}\right\}, \tag{3}$$

where r denotes electrode resistances, μ the mu-factors, and S the transconductances.

It follows readily from (2) and (3) that

$$\left.\begin{aligned}
b_{10} &= \frac{1}{r_p} & \beta_{10} &= \frac{\mu_g}{r_g} \\
b_{01} &= \frac{\mu_p}{r_p} & \beta_{01} &= \frac{1}{r_g} \\
b_{20} &= -\frac{1}{2r_p^2}\frac{\partial r_p}{\partial E_p} = P_2 & \beta_{02} &= -\frac{1}{2r_g^2}\frac{\partial r_g}{\partial E_g} = T_2 \\
b_{11} &= \frac{1}{r_p}\frac{\partial \mu_p}{\partial E_p} + 2\mu_p P_2 & \beta_{11} &= \frac{1}{r_g}\frac{\partial \mu_g}{\partial E_g} + 2\mu_g T_2 \\
b_{02} &= \frac{1}{2r_p}\frac{\partial \mu_p}{\partial E_g} + \frac{\mu_p}{2r_p}\frac{\partial \mu_p}{\partial E_p} + \mu_p^2 P_2 & \beta_{20} &= \frac{1}{2r_g}\frac{\partial \mu_g}{\partial E_p} + \frac{\mu_g}{2r_g}\frac{\partial \mu_g}{\partial E_g} + \mu_g^2 T_2
\end{aligned}\right\}, \tag{4}$$

where P_2 and T_2 are new notations for b_{20} and β_{02}, respectively. Similar expressions may be derived for the coefficients b_{30}, β_{30}, etc.

If we now apply the circuital laws to the network external to the tube, we get a number of equations, two of which are

$$\epsilon_g = g + e_g, \qquad \epsilon_p = p + e_p. \tag{5}$$

To obtain a solution of (1) and (5) we utilize a method of successive approximations. Let

$$i_p = \sum i_{pk},\ i_g = \sum i_{gk},\ e_p = \sum e_{pk},\ e_g = \sum e_{gk},\ g = \sum g_k,\ p = \sum p_k, \tag{6}$$

where the summations extend from $k = 1$ to $k = \infty$. Let us further define the terms in the series (6) by the following equations:

$$\left.\begin{aligned}
r_p i_{p1} - e_{p1} &= \mu_p e_{g1} \\
r_g i_{g1} - e_{g1} &= \mu_g e_{p1} \\
\epsilon_g = g_1 + e_{g1}, \qquad \epsilon_p &= p_1 + e_{p1}
\end{aligned}\right\}, \tag{7}$$

$$\left.\begin{aligned}
r_p i_{p2} - e_{p2} &= \mu_p e_{g2} + r_p(b_{20}e_{p1}^2 + b_{11}e_{p1}e_{g1} + b_{02}e_{g1}^2) \\
r_g i_{g2} - e_{g2} &= \mu_g e_{p2} + r_g(\beta_{20}e_{p1}^2 + \beta_{11}e_{p1}e_{g1} + \beta_{02}e_{g1}^2) \\
0 &= g_2 + e_{g2}, \qquad 0 = p_2 + e_{p2}
\end{aligned}\right\}, \tag{8}$$

$$\left.\begin{aligned}
r_p i_{p3} - e_{p3} &= \mu_p e_{g3} + r_p[2b_{20}e_{p1}e_{p2} + b_{11}(e_{p1}e_{g2} + e_{p2}e_{g1}) + 2b_{02}e_{g1}e_{g2} \\
&\qquad + b_{30}e_{p1}^3 + b_{21}e_{p1}^2e_{g1} + b_{12}e_{p1}e_{g1}^2 + b_{03}e_{g1}^3] \\
r_g i_{g3} - e_{g3} &= \mu_g e_{p3} + r_g[2\beta_{20}e_{p1}e_{p2} + \beta_{11}(e_{p1}e_{g2} + e_{p2}e_{g1}) + 2\beta_{02}e_{g1}e_{g2} \\
&\qquad + \beta_{30}e_{p1}^3 + \beta_{21}e_{p1}^2e_{g1} + \beta_{12}e_{p1}e_{g1}^2 + \beta_{03}e_{g1}^3] \\
0 &= g_3 + e_{g3}, \qquad 0 = p_3 + e_{p3}
\end{aligned}\right\} \tag{9}$$

and so forth for subsequent terms.[5]

The physical interpretation of equations (7) to (9) is readily obtained. It follows from (7) that the equivalent circuit of Fig. 1 for first order quantities is given by Fig. 2. The equivalent circuit of

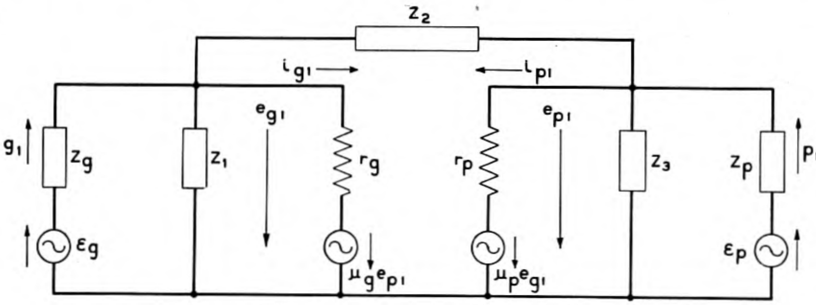


Fig. 2—Equivalent circuit—first-order effects.

Fig. 1 for second and third order effects are those shown in Fig. 3, and Fig. 4, respectively.

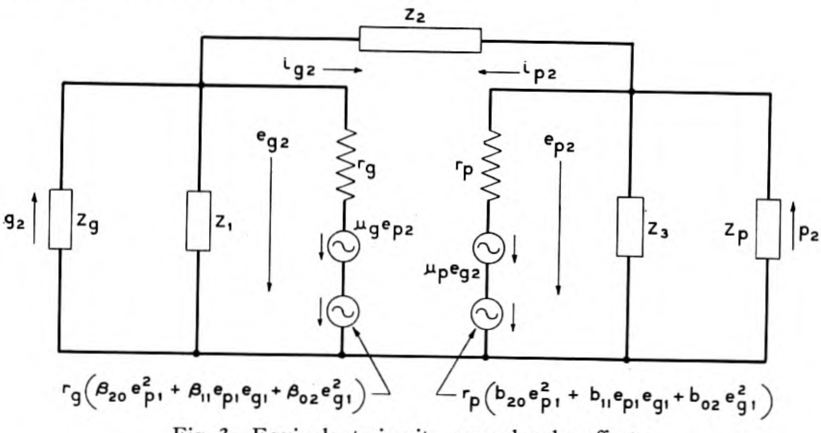


Fig. 3—Equivalent circuit—second-order effects.

It follows from (4) and (8) that

$$
\left\{
\begin{aligned}
& r_p(b_{20}e_{p1}^2 + b_{11}e_{p1}e_{g1} + b_{02}e_{g1}^2) \\
& \qquad = r_p P_2(e_{p1} + \mu_p e_{g1})^2 + \frac{1}{2}\left(\frac{\partial \mu_p}{\partial E_g} + \mu_p \frac{\partial \mu_p}{\partial E_p}\right) e_{g1}^2 + \frac{\partial \mu_p}{\partial E_p} e_{p1}e_{g1}, \\
& r_g(\beta_{20}e_{p1}^2 + \beta_{11}e_{p1}e_{g1} + \beta_{02}e_{g1}^2) \\
& \qquad = r_g T_2(e_{g1} + \mu_g e_{p1})^2 + \frac{1}{2}\left(\frac{\partial \mu_g}{\partial E_p} + \mu_g \frac{\partial \mu_g}{\partial E_g}\right) e_{p1}^2 + \frac{\partial \mu_g}{\partial E_g} e_{p1}e_{g1}.
\end{aligned}
\right. \tag{10}
$$

The corresponding terms in (9) can be expressed similarly.

Equations (7), (8) and (9) contain the general theory of the three-electrode vacuum tube circuit. In the special case when conductive grid current is absent it is only necessary to omit the second equation

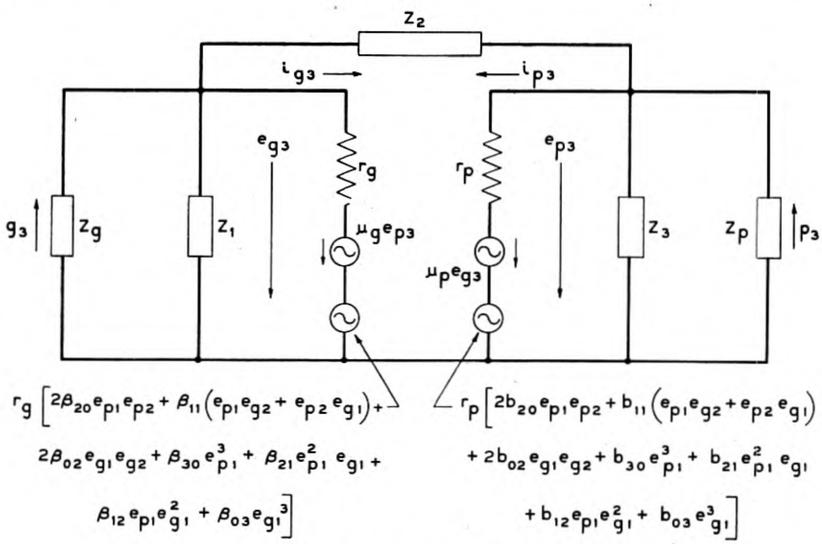


Fig. 4—Equivalent circuit—third-order effects.

in each of the equations (7) to (10), inclusive, and to omit in each of the Figs. 2, 3, and 4, the branch containing r_g. If it is assumed that not only conductive grid current is absent, but also that μ_p is constant, the second plate e.m.f. in (8) reduces to $r_pP_2(e_{p1} + \mu_p e_{g1})^2$ as is seen from (10), and (8) thus becomes identical with the corresponding equation already obtained previously.[5] A similar reduction and correspondence occurs for the second plate e.m.f. in (9), as well as in subsequent equations.

Application to Steady State Solutions

In this section the use of the theory is illustrated by obtaining first and second-order effects assuming the circuit configuration to be that shown in Fig. 1. To avoid unnecessary complications the discussion is limited to steady-state solutions, and it is also assumed that no plate e.m.f. is impressed. We shall first obtain the solutions in the general case and then indicate how these are simplified in such special cases which have been treated by some previous investigators.[2, 3, 5]

General Case

Let the impressed grid e.m.f. be

$$\epsilon_g = \sum k_h \cos(\omega_h t + \kappa_h) = R\sum k_h e^{i(\omega_h t + \kappa_h)}, \qquad i = \sqrt{-1} \quad (11)$$

where the summation extends from $h = 1$ to $h = n$, and the letter R before an expression means its real part. Referring to Fig. 2 it may be shown that

$$\left.\begin{aligned} e_{g1} &= R\sum \frac{\alpha_1(\omega_h)}{Z(\omega_h)} k_h e^{i(\omega_h t+\kappa_h)} \\ e_{p1} &= R\sum -\frac{\alpha_2(\omega_h)}{Z(\omega_h)} k_h e^{i(\omega_h t+\kappa_h)} \\ (e_{p1} + \mu_p e_{g1}) &= R\sum r_p \frac{\alpha_3(\omega_h)}{Z(\omega_h)} k_h e^{i(\omega_h t+\kappa_h)} \\ (e_{g1} + \mu_g e_{p1}) &= R\sum \frac{\alpha_4(\omega_h)}{Z(\omega_h)} k_h e^{i(\omega_h t+\kappa_h)} \end{aligned}\right\}, \quad (12)$$

where

$$\begin{aligned} Z(\omega) = \frac{1}{(Z_g' + Z_2)r_g Z_2} \{ &[Z_2(r_g + Z_g') + r_g Z_g'] \\ &\times [Z_2(r_p + Z_p') + r_p Z_p'] \\ &- Z_g' Z_p' (\mu_g Z_2 - r_g)(\mu_p Z_2 - r_p)\}, \end{aligned} \quad (13)$$

$$\alpha_1(\omega) = \frac{Z_1[Z_2(r_p + Z_p') + r_p Z_p']}{(Z_1 + Z_g)(Z_2 + Z_g')}, \quad (14)$$

$$\alpha_2(\omega) = \frac{Z_1 Z_p'(\mu_p Z_2 - r_p)}{(Z_1 + Z_g)(Z_2 + Z_g')}, \quad (15)$$

$$\alpha_3(\omega) = \frac{Z_1(\mu_p Z_2 + \mu_p Z_p' + Z_p')}{(Z_1 + Z_g)(Z_2 + Z_g')}, \quad (16)$$

$$\alpha_4(\omega) = \frac{Z_1 Z_2 r_p + Z_1 Z_2 Z_p'(1 - \mu_g \mu_p) + Z_1 r_p Z_p'(1 + \mu_g)}{(Z_1 + Z_g)(Z_2 + Z_g')}, \quad (17)$$

$$Z_p'(\omega) = \frac{Z_3 Z_p}{Z_3 + Z_p}, \qquad Z_g'(\omega) = \frac{Z_1 Z_g}{Z_1 + Z_g}. \quad (18)$$

The right-hand expressions in (13)–(18) are to be evaluated at ω. If we write

$$\frac{\alpha_k(\omega)}{Z(\omega)} = \left|\frac{\alpha_k(\omega)}{Z(\omega)}\right| e^{-i\varphi_k(\omega)}, \qquad k = 1, 2, 3, 4, \quad (19)$$

the equations (12) can be written

$$\left.\begin{aligned} e_{g1} &= \sum \left| \frac{\alpha_1(\omega_h)}{Z(\omega_h)} \right| k_h \cos [\omega_h t + \kappa_h - \varphi_1(\omega_h)] \\ e_{p1} &= -\sum \left| \frac{\alpha_2(\omega_h)}{Z(\omega_h)} \right| k_h \cos [\omega_h t + \kappa_h - \varphi_2(\omega_h)] \\ e_{p1} + \mu_p e_{g1} &= \sum r_p \left| \frac{\alpha_3(\omega_h)}{Z(\omega_h)} \right| k_h \cos [\omega_h t + \kappa_h - \varphi_3(\omega_h)] \\ e_{g1} + \mu_g e_{p1} &= \sum \left| \frac{\alpha_4(\omega_h)}{Z(\omega_h)} \right| k_h \cos [\omega_h t + \kappa_h - \varphi_4(\omega_h)] \end{aligned}\right\}. \tag{20}$$

This concludes our consideration of effects of the first order and we now turn to those of the second order. For this purpose we substitute the values given by (20) in the right-hand side of the expressions (10) for the grid and plate e.m.f.'s, and we then obtain two expressions, each of which is equal to a sum of sinusoidal terms. If we limit our attention to the terms of frequency $(\omega_1 - \omega_2)$, it is readily shown that the grid e.m.f. of this frequency is equal to the real part of

$$\begin{aligned} &\left[r_g T_2 \frac{\alpha_4(\omega_1)}{Z(\omega_1)} \left(\overline{\frac{\alpha_4(\omega_2)}{Z(\omega_2)}} \right) + \frac{1}{2} \left(\frac{\partial \mu_g}{\partial E_p} + \mu_g \frac{\partial \mu_g}{\partial E_g} \right) \frac{\alpha_2(\omega_1)}{Z(\omega_1)} \left(\overline{\frac{\alpha_2(\omega_2)}{Z(\omega_2)}} \right) \right. \\ &\left. - \frac{1}{2} \frac{\partial \mu_g}{\partial E_g} \left\{ \frac{\alpha_1(\omega_1)}{Z(\omega_1)} \left(\overline{\frac{\alpha_2(\omega_2)}{Z(\omega_2)}} \right) + \left(\overline{\frac{\alpha_1(\omega_2)}{Z(\omega_2)}} \right) \frac{\alpha_2(\omega_1)}{Z(\omega_1)} \right\} \right] k_1 k_2 e^{i((\omega_1 - \omega_2)t + \kappa_1 - \kappa_2)} \end{aligned} \tag{21}$$

and the plate e.m.f. is equal to the real part of

$$\begin{aligned} &\left[r_p{}^3 P_2 \frac{\alpha_3(\omega_1)}{Z(\omega_1)} \left(\overline{\frac{\alpha_3(\omega_2)}{Z(\omega_2)}} \right) + \frac{1}{2} \left(\frac{\partial \mu_p}{\partial E_g} + \mu_p \frac{\partial \mu_p}{\partial E_p} \right) \frac{\alpha_1(\omega_1)}{Z(\omega_1)} \left(\overline{\frac{\alpha_1(\omega_2)}{Z(\omega_2)}} \right) \right. \\ &\left. - \frac{1}{2} \frac{\partial \mu_p}{\partial E_p} \left\{ \frac{\alpha_1(\omega_1)}{Z(\omega_1)} \left(\overline{\frac{\alpha_2(\omega_2)}{Z(\omega_2)}} \right) + \left(\overline{\frac{\alpha_1(\omega_2)}{Z(\omega_2)}} \right) \frac{\alpha_2(\omega_1)}{Z(\omega_1)} \right\} \right] k_1 k_2 e^{i((\omega_1 - \omega_2)t + \kappa_1 - \kappa_2)}, \end{aligned} \tag{22}$$

where a bar over a quotient indicates its conjugate complex.

It follows from Fig. 3 by straightforward calculations that the currents i_{p2} and i_{g2} produced by the e.m.f.'s (21) and (22), are

$$\left.\begin{aligned} i_{p2}(\omega_1 - \omega_2) &= R \left[-\frac{[\epsilon_g]}{Z_b(\omega_1 - \omega_2)} + \frac{[\epsilon_p]}{Z_d(\omega_1 - \omega_2)} \right] \\ i_{g2}(\omega_1 - \omega_2) &= R \left[\frac{[\epsilon_g]}{Z_a(\omega_1 - \omega_2)} - \frac{[\epsilon_p]}{Z_c(\omega_1 - \omega_2)} \right] \end{aligned}\right\}, \tag{23}$$

where $[\epsilon_g]$ and $[\epsilon_p]$ are abbreviations for the complex quantities (21) and (22), respectively, and

$$\left.\begin{aligned} Z_a(\omega) &= Z(\omega)\frac{r_g(Z_g' + Z_2)}{Z_2(r_p + Z_p') + r_pZ_p'} = |Z_a(\omega)|e^{i\psi_a(\omega)} \\ Z_b(\omega) &= Z(\omega)\frac{r_g(Z_g' + Z_2)}{Z_g'(\mu_pZ_2 - r_p)} = |Z_b(\omega)|e^{i\psi_b(\omega)} \\ Z_c(\omega) &= Z(\omega)\frac{r_g(Z_2 + Z_g')}{Z_p'(\mu_gZ_2 - r_g)} = |Z_c(\omega)|e^{i\psi_c(\omega)} \\ Z_d(\omega) &= Z(\omega)\frac{r_g(Z_2 + Z_g')}{Z_2(r_g + Z_g') + r_gZ_g'} = |Z_d(\omega)|e^{i\psi_d(\omega)} \end{aligned}\right\}. \tag{24}$$

In (24) the introduction of the angles ψ is convenient when it is desired to evaluate the real parts of the expressions (23).

The expressions (21) to (24) can be used to obtain any second-order current of frequency $(\omega_a - \omega_b)$ by replacing ω_1 with ω_a and ω_2 with ω_b. The remaining second-order currents are found by a process similar to that above. For instance, $i_{p2}(2\omega_1)$ and $i_{g2}(2\omega_1)$ are given by the right-hand expressions in (23) provided the e.m.f.'s $[\epsilon]$ are those of frequency $(2\omega_1)$ and the impedances Z are evaluated at $(2\omega_1)$. In passing it may be remarked that equations similar to those in (23) and (24) also occur when third and higher-order effects are calculated.

Special Cases

If the impedances Z_1, Z_2, and Z_3 are infinite the case treated above reduces to that considered by Llewellyn,[3] and after proper simplifications the previous equations give results identical with those obtained by him. For instance, if we take the limiting values of e_{p1} in (12) as Z_1, Z_2, and Z_3 tend to infinity, and if we then divide the quantity inside the summation sign by $-Z_p(\omega_h)$, we get an expression for i_{p1} which may be shown to be identical with equation (33) in Llewellyn's paper, except for differences in notations. Similarly, the plate current $i_{p2}(\omega_1 - \omega_2)$ in (23) reduces to a value which may be shown to be equal to the sum of his equations (35) and (36), evaluated for this type of second-order current.

Another special case is that when the impedances Z_1, Z_2, and Z_3 are all finite but conductive grid current is absent. We then have μ_g equal to zero, and R_g equal to infinity, and the previous general equations are simplified correspondingly.

We arrive at the case treated by Peterson-Evans [2] by maintaining the assumption of no conductive grid current but by assuming Z_1, Z_2, and Z_3 to be infinite. For instance, if then i_{p1} and $i_{p2}(\omega_1 - \omega_2)$ are evaluated on this basis for a plate impedance Z_p equal to a pure resistance at all frequencies, it can be shown that the currents so obtained are identical with the corresponding currents given by equations (4) and (6) in the paper referred to.

Finally, if we assume finite values for Z_1, Z_2, and Z_3, no conductive grid current, and constant μ_p, we have the case treated in the previous paper.[5]

Theory of Four-Electrode Tube Circuits

Circuits with tubes having more than three electrodes can be treated by a process similar to that adopted above, as will be made clear by outlining the theory for the four-electrode tube circuit.

The circuit to be considered is shown in Fig. 5 where Z_1 to Z_6 are

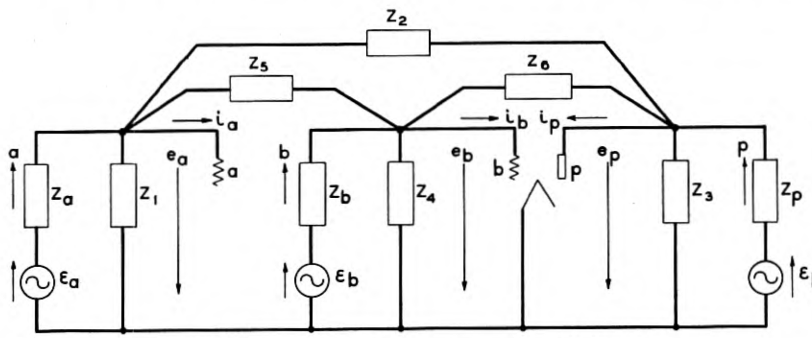


Fig. 5—Four-electrode vacuum tube and circuit.

impedances which may include inter-electrode admittances. On the electrodes denoted by a, b, and p are impressed the variable electromotive forces ϵ_a, ϵ_b, and ϵ_p in series with the impedances Z_a, Z_b, and Z_p, respectively. The significance of the quantities E_p, E_{p0}, e_p, I_p, I_{p0}, i_p and the corresponding quantities with indices a and b, is obvious from the preceding discussion of the three-electrode tube circuit. Let a, b, and p be the incremental voltages across the impedances Z_a, Z_b, and Z_p. As before, instantaneous values are implied.

For the currents we get the expansions

$$\left.\begin{aligned} i_p &= P_1e_a + P_2e_b + P_3e_p + P_4e_a^2 + P_5e_b^2 + P_6e_p^2 \\ &\qquad + P_7e_ae_b + P_8e_ae_p + P_9e_be_p + \cdots \\ i_a &= A_1e_a + A_2e_b + A_3e_p + A_4e_a^2 + A_5e_b^2 + A_6e_p^2 \\ &\qquad + A_7e_ae_b + A_8e_ae_p + A_9e_be_p + \cdots \\ i_b &= B_1e_a + B_2e_b + B_3e_p + B_4e_a^2 + B_5e_b^2 + B_6e_p^2 \\ &\qquad + B_7e_ae_b + B_8e_ae_p + B_9e_be_p + \cdots \end{aligned}\right\}, \tag{25}$$

where

$$\left.\begin{aligned} P_1 &= \frac{\partial I_p}{\partial E_a},\ P_2 = \frac{\partial I_p}{\partial E_b},\ P_3 = \frac{\partial I_p}{\partial E_p},\ P_4 = \frac{1}{2}\frac{\partial^2 I_p}{\partial E_a^2},\ P_5 = \frac{1}{2}\frac{\partial^2 I_p}{\partial E_b^2} \\ P_6 &= \frac{1}{2}\frac{\partial^2 I_p}{\partial E_p^2},\ P_7 = \frac{\partial^2 I_p}{\partial E_a \partial E_b},\ P_8 = \frac{\partial^2 I_p}{\partial E_a \partial E_p},\ P_9 = \frac{\partial^2 I_p}{\partial E_b \partial E_p},\ \cdots \end{aligned}\right\} \tag{26}$$

and similar expressions hold for the A- and B-values.

The electrode resistances are by definition

$$\frac{1}{r_p} = \frac{\partial I_p}{\partial E_p}, \qquad \frac{1}{r_a} = \frac{\partial I_a}{\partial E_a}, \qquad \frac{1}{r_b} = \frac{\partial I_b}{\partial E_b}. \tag{27}$$

The mu-factors are

$$\left.\begin{aligned} \mu_{pa} &= \frac{\dfrac{\partial I_p}{\partial E_a}}{\dfrac{\partial I_p}{\partial E_p}} = -\left(\frac{dE_p}{dE_a}\right)_{E_b,\ I_p=\text{const.}} \\ \mu_{ap} &= \frac{\dfrac{\partial I_a}{\partial E_p}}{\dfrac{\partial I_a}{\partial E_a}} = -\left(\frac{dE_a}{dE_p}\right)_{E_b,\ I_a=\text{const.}} \\ \mu_{bp} &= \frac{\dfrac{\partial I_b}{\partial E_p}}{\dfrac{\partial I_b}{\partial E_b}} = -\left(\frac{dE_b}{dE_p}\right)_{E_a,\ I_b=\text{const.}} \\ \mu_{pb} &= \frac{\dfrac{\partial I_p}{\partial E_b}}{\dfrac{\partial I_p}{\partial E_p}} = -\left(\frac{dE_p}{dE_b}\right)_{E_a,\ I_p=\text{const.}} \\ \mu_{ab} &= \frac{\dfrac{\partial I_a}{\partial E_b}}{\dfrac{\partial I_a}{\partial E_a}} = -\left(\frac{dE_a}{dE_b}\right)_{E_p,\ I_a=\text{const.}} \\ \mu_{ba} &= \frac{\dfrac{\partial I_b}{\partial E_a}}{\dfrac{\partial I_b}{\partial E_b}} = -\left(\frac{dE_b}{dE_a}\right)_{E_p,\ I_b=\text{const.}} \end{aligned}\right\} \tag{28}$$

and the transconductances

$$\left.\begin{aligned} S_{pa} &= \frac{\partial I_p}{\partial E_a} = \frac{\mu_{pa}}{r_p} & S_{pb} &= \frac{\partial I_p}{\partial E_b} = \frac{\mu_{pb}}{r_p} \\ S_{ap} &= \frac{\partial I_a}{\partial E_p} = \frac{\mu_{ap}}{r_a} & S_{ab} &= \frac{\partial I_a}{\partial E_b} = \frac{\mu_{ab}}{r_a} \\ S_{bp} &= \frac{\partial I_b}{\partial E_p} = \frac{\mu_{bp}}{r_b} & S_{ba} &= \frac{\partial I_b}{\partial E_a} = \frac{\mu_{ba}}{r_b} \end{aligned}\right\}. \tag{29}$$

It can now be shown that

$$\left.\begin{aligned} P_1 &= \frac{\mu_{pa}}{r_p}, \qquad P_2 = \frac{\mu_{pb}}{r_p}, \qquad P_3 = \frac{1}{r_p} \\ P_6 &= -\frac{1}{2r_p^2}\frac{\partial r_p}{\partial E_p}, \; P_8 = \frac{1}{r_p}\frac{\partial \mu_{pa}}{\partial E_p} + 2\mu_{pa}P_6, \; P_9 = \frac{1}{r_p}\frac{\partial \mu_{pb}}{\partial E_p} + 2\mu_{pb}P_6 \\ P_4 &= \frac{1}{2r_p}\frac{\partial \mu_{pa}}{\partial E_a} + \frac{\mu_{pa}}{2r_p}\frac{\partial \mu_{pa}}{\partial E_p} + \mu_{pa}^2 P_6 \\ P_5 &= \frac{1}{2r_p}\frac{\partial \mu_{pb}}{\partial E_b} + \frac{\mu_{pb}}{2r_p}\frac{\partial \mu_{pb}}{\partial E_p} + \mu_{pb}^2 P_6 \\ P_7 &= \frac{1}{r_p}\frac{\partial \mu_{pb}}{\partial E_a} + \frac{\mu_{pb}}{r_p}\frac{\partial \mu_{pa}}{\partial E_p} + 2\mu_{pb}\mu_{pa}P_6 \\ &= \frac{1}{r_p}\frac{\partial \mu_{pa}}{\partial E_b} + \frac{\mu_{pa}}{r_p}\frac{\partial \mu_{pb}}{\partial E_p} + 2\mu_{pa}\mu_{pb}P_6 \end{aligned}\right\}, \tag{30}$$

$$\left.\begin{aligned} A_1 &= \frac{1}{r_a}, \qquad A_2 = \frac{\mu_{ab}}{r_a}, \qquad A_3 = \frac{\mu_{ap}}{r_a} \\ A_4 &= -\frac{1}{2r_a^2}\frac{\partial r_a}{\partial E_a}, \; A_7 = \frac{1}{r_a}\frac{\partial \mu_{ab}}{\partial E_a} + 2\mu_{ab}A_4, \; A_8 = \frac{1}{r_a}\frac{\partial \mu_{ap}}{\partial E_a} + 2\mu_{ap}A_4 \\ A_5 &= \frac{1}{2r_a}\frac{\partial \mu_{ab}}{\partial E_b} + \frac{\mu_{ab}}{2r_a}\frac{\partial \mu_{ab}}{\partial E_a} + \mu_{ab}^2 A_4 \\ A_6 &= \frac{1}{2r_a}\frac{\partial \mu_{ap}}{\partial E_p} + \frac{\mu_{ap}}{2r_a}\frac{\partial \mu_{ap}}{\partial E_a} + \mu_{ap}^2 A_4 \\ A_9 &= \frac{1}{r_a}\frac{\partial \mu_{ap}}{\partial E_b} + \frac{\mu_{ap}}{r_a}\frac{\partial \mu_{ab}}{\partial E_a} + 2\mu_{ap}\mu_{ab}A_4 \\ &= \frac{1}{r_a}\frac{\partial \mu_{ab}}{\partial E_p} + \frac{\mu_{ab}}{r_a}\frac{\partial \mu_{ap}}{\partial E_a} + 2\mu_{ab}\mu_{ap}A_4 \end{aligned}\right\}, \tag{31}$$

$$\left.\begin{aligned}
B_1 &= \frac{\mu_{ba}}{r_b}, \qquad B_2 = \frac{1}{r_b}, \qquad B_3 = \frac{\mu_{bp}}{r_b} \\
B_5 &= -\frac{1}{2r_b^2}\frac{\partial r_b}{\partial E_b}, \; B_7 = \frac{1}{r_b}\frac{\partial \mu_{ba}}{\partial E_b} + 2\mu_{ba}B_5, \; B_9 = \frac{1}{r_b}\frac{\partial \mu_{bp}}{\partial E_b} + 2\mu_{bp}B_5 \\
B_4 &= \frac{1}{2r_b}\frac{\partial \mu_{ba}}{\partial E_a} + \frac{\mu_{ba}}{2r_b}\frac{\partial \mu_{ba}}{\partial E_b} + \mu_{ba}^2 B_5 \\
B_6 &= \frac{1}{2r_b}\frac{\partial \mu_{bp}}{\partial E_p} + \frac{\mu_{bp}}{2r_b}\frac{\partial \mu_{bp}}{\partial E_b} + \mu_{bp}^2 B_5 \\
B_8 &= \frac{1}{r_b}\frac{\partial \mu_{bp}}{\partial E_a} + \frac{\mu_{bp}}{r_b}\frac{\partial \mu_{ba}}{\partial E_b} + 2\mu_{bp}\mu_{ba}B_5 \\
&= \frac{1}{r_b}\frac{\partial \mu_{ba}}{\partial E_p} + \frac{\mu_{ba}}{r_b}\frac{\partial \mu_{bp}}{\partial E_p} + 2\mu_{ba}\mu_{bp}B_5
\end{aligned}\right\}. \tag{32}$$

The circuital laws applied to the external network furnish a number of equations, three of which are

$$\epsilon_a = a + e_a, \qquad \epsilon_b = b + e_b, \qquad \epsilon_p = p + e_p. \tag{33}$$

Let now

$$\left.\begin{aligned}
i_p &= \sum i_{pk}, & i_a &= \sum i_{ak}, & i_b &= \sum i_{bk} \\
e_p &= \sum e_{pk}, & e_a &= \sum e_{ak}, & e_b &= \sum e_{bk} \\
p &= \sum p_k, & a &= \sum a_k, & b &= \sum b_k
\end{aligned}\right\}. \tag{34}$$

We then obtain the equations

$$\left.\begin{aligned}
r_p i_{p1} - e_{p1} &= \mu_{pa}e_{a1} + \mu_{pb}e_{b1} \\
r_a i_{a1} - e_{a1} &= \mu_{ab}e_{b1} + \mu_{ap}e_{p1} \\
r_b i_{b1} - e_{b1} &= \mu_{ba}e_{a1} + \mu_{bp}e_{p1} \\
\epsilon_a = a_1 + e_{a1}, \quad \epsilon_b &= b_1 + e_{b1}, \quad \epsilon_p = p_1 + e_{p1}
\end{aligned}\right\}, \tag{35}$$

which show that the equivalent circuit for first order effects is that given in Fig. 6.

We get further for second-order quantities

$$\left.\begin{aligned}
r_p i_{p2} - e_{p2} &= \mu_{pa}e_{a2} + \mu_{pb}e_{b2} + r_p L \\
r_a i_{a2} - e_{a2} &= \mu_{ab}e_{b2} + \mu_{ap}e_{p2} + r_a M \\
r_b i_{b2} - e_{b2} &= \mu_{ba}e_{a2} + \mu_{bp}e_{p2} + r_b N \\
0 = a_2 + e_{a2}, \quad 0 &= b_2 + e_{b2}, \quad 0 = p_2 + e_{p2}
\end{aligned}\right\}, \tag{36}$$

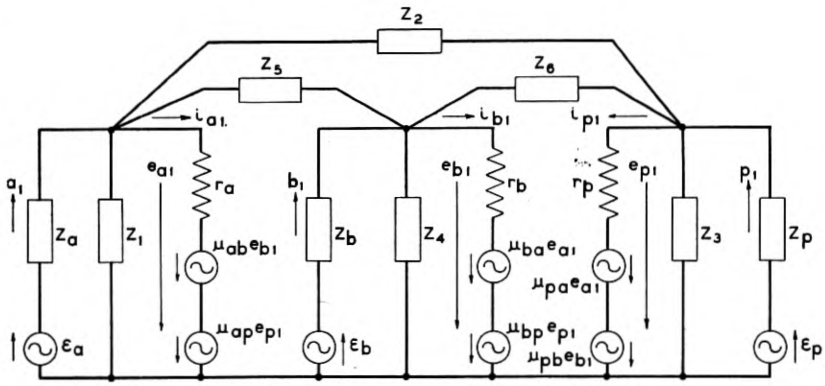


Fig. 6—Equivalent circuit—first-order effects.

where

$$\left.\begin{aligned} r_pL &= (P_4e_{a1}^2+P_5e_{b1}^2+P_6e_{p1}^2+P_7e_{a1}e_{b1}+P_8e_{a1}e_{p1}+P_9e_{b1}e_{p1})r_p \\ r_aM &= (A_4e_{a1}^2+A_5e_{b1}^2+A_6e_{p1}^2+A_7e_{a1}e_{b1}+A_8e_{a1}e_{p1}+A_9e_{b1}e_{p1})r_a \\ r_bN &= (B_4e_{a1}+B_5e_{b1}+B_6e_{p1}+B_7e_{a1}e_{b1}+B_8e_{a1}e_{p1}+B_9e_{b1}e_{p1})r_b \end{aligned}\right\}, \quad (37)$$

which in view of (30), (31) and (32) may be written

$$\left.\begin{aligned} r_pL = {} & r_pP_6(\mu_{pa}e_{a1} + \mu_{pb}e_{b1} + e_{p1})^2 + \frac{1}{2}\left(\frac{\partial\mu_{pa}}{\partial E_a} + \mu_{pa}\frac{\partial\mu_{pa}}{\partial E_p}\right)e_{a1}^2 \\ & + \frac{1}{2}\left(\frac{\partial\mu_{pb}}{\partial E_b} + \mu_{pb}\frac{\partial\mu_{pb}}{\partial E_p}\right)e_{b1}^2 + \left(\frac{\partial\mu_{pb}}{\partial E_a} + \mu_{pb}\frac{\partial\mu_{pa}}{\partial E_p}\right)e_{a1}e_{b1} \\ & + \frac{\partial\mu_{pa}}{\partial E_p}e_{a1}e_{p1} + \frac{\partial\mu_{pb}}{\partial E_p}e_{b1}e_{p1} \end{aligned}\right\}, \quad (38)$$

$$\left.\begin{aligned} r_aM = {} & r_aA_4(e_{a1} + \mu_{ab}e_{b1} + \mu_{ap}e_{p1})^2 + \frac{1}{2}\left(\frac{\partial\mu_{ab}}{\partial E_b} + \mu_{ab}\frac{\partial\mu_{ab}}{\partial E_a}\right)e_{b1}^2 \\ & + \frac{1}{2}\left(\frac{\partial\mu_{ap}}{\partial E_p} + \mu_{ap}\frac{\partial\mu_{ap}}{\partial E_a}\right)e_{p1}^2 + \frac{\partial\mu_{ab}}{\partial E_a}e_{a1}e_{b1} \\ & + \frac{\partial\mu_{ap}}{\partial E_a}e_{a1}e_{p1} + \left(\frac{\partial\mu_{ap}}{\partial E_b} + \mu_{ap}\frac{\partial\mu_{ab}}{\partial E_a}\right)e_{b1}e_{p1} \end{aligned}\right\}, \quad (39)$$

$$\left.\begin{aligned} r_bN = {} & r_bB_5(\mu_{ba}e_{a1} + e_{b1} + \mu_{bp}e_{p1})^2 + \frac{1}{2}\left(\frac{\partial\mu_{ba}}{\partial E_a} + \mu_{ba}\frac{\partial\mu_{ba}}{\partial E_b}\right)e_{a1}^2 \\ & + \frac{1}{2}\left(\frac{\partial\mu_{bp}}{\partial E_p} + \mu_{bp}\frac{\partial\mu_{bp}}{\partial E_b}\right)e_{p1}^2 + \frac{\partial\mu_{ba}}{\partial E_b}e_{a1}e_{b1} \\ & + \left(\frac{\partial\mu_{bp}}{\partial E_a} + \mu_{bp}\frac{\partial\mu_{ba}}{\partial E_b}\right)e_{a1}e_{p1} + \frac{\partial\mu_{bp}}{\partial E_b}e_{b1}e_{p1} \end{aligned}\right\}. \quad (40)$$

From (36) it now follows that the equivalent circuit for second-order effects for the four-electrode tube circuit is as indicated in Fig. 7.

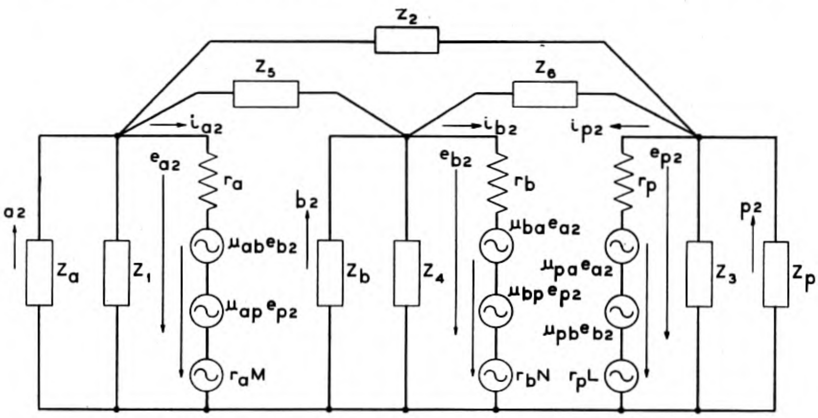


Fig. 7—Equivalent circuit—second-order effects.

Transatlantic Long-Wave Radio Telephone Transmission and Related Phenomena from 1923 to 1933

By AUSTIN BAILEY and HOWARD M. THOMSON

It is shown that transatlantic long-wave radio field strength is related to the 11-year cycles of terrestrial magnetic activity, sunspots, solar limb-prominences, and ultra-violet radiation. The directness of correlation between long-wave radio and these other phenomena is apparently only approximate. Very good seasonal and monthly correlation is obtained between magnetic activity and daylight radio transmission. Magnetic storms are shown to have prolonged and delayed effects on day and night radio transmission, obscuring tendencies for 27-day recurrences on long waves. No reference is made to the probable mechanism of long-wave radio transmission because a paper, now in preparation, will be concerned primarily with this subject.

Introduction

THIS paper summarizes a study of the transatlantic long-wave radio transmission data (near 60 kilocycles) collected by the American Telephone and Telegraph Company and the General Post Office of Great Britain during the period from 1923 to 1933, including the development tests from 1923 to 1926 and the period of operation of the commercial radio telephone circuits from January 1927 to December 1933.

A correlation is presented between terrestrial magnetic activity, sunspot-numbers, solar limb-prominences, ultra-violet radiation, and transatlantic long-wave radio telephone field strength observations during one 11-year sunspot cycle. By expressing the variables with different scales better correlations have been obtained. These are given on averages of years, months, seasons, and days. Examples of delayed night and day field strength changes accompanying long magnetic storms are included in this presentation.

It is generally believed that solar radiation influences the transmission [1*] of radio signals but the detailed mechanism by which the influence is exerted [2] is not entirely clear. The number and relative effects of other factors that may influence radio transmission are also uncertain. Solar phenomena, magnetic activity, and radio transmission appear to have some time-phase relations but these relations have not been clearly established. Some of the results of the studies reported herein contribute to the evaluation of these effects while other results give new or corroborative information on the qualitative relations between the solar, magnetic, and radio phenomena.

* Numbers refer to the references in the bibliography.

In any particular study, the field strengths of the radio transmitters were corrected to a constant antenna current and expressed in decibels above one microvolt per meter. Where a study involved a comparison between data observed at different locations, or on different frequencies, corrections were applied to reduce the data to a common basis by the application of the Bell System Long-Wave Radio Transmission Formula.[3]

The transmitters and transmission paths involved are indicated in Table I.

TABLE I

Years	Transmitter		Rocky Point, N. Y., to
	West–East		
1923–1924	57 Kc.	2XS	New Southgate, England
1925	57 Kc.	2XS	Chedzoy, England
1926 (to Sept.)	57 Kc.	2XS	Wroughton, England
1926 (Oct.) to 1928	60 Kc.	WNL	Wroughton, England
1927 (Mar.) to 1933	60 Kc.	WNL	Cupar, Scotland
	East–West		Rugby, England, to
1927–1930, part of 1931	60 Kc.	GBT	Houlton, Maine
Remainder of 1931	68 Kc.	GKA	Houlton, Maine
1932–1933	68 Kc.	GBY	Houlton, Maine

Statistical Frequency Distributions

In making a correlation study the best results are obtained when observations of the variables involved have the same type of statistical "frequency distribution." As illustrated in Fig. 1, the cumulative frequency distribution curves of long-wave field strength observations, when plotted (in decibels) on "arithmetic probability" paper, tend to be straight lines and thus of the Normal Law type. The best correlation will therefore be obtained when the other variables involved have approximately straight-line distributions on the same type (arithmetic probability) of coordinates.

Accordingly, cumulative distribution curves were so made for all the variables used in this study. It was found that good approximations to the Normal Law distribution were obtained with logarithmic scales for radio field strength (db above 1 $\mu v/m$), monthly averages of solar limb-prominences, sunspot-numbers, and u_1 measure * of magnetic activity. However, for the "C" measure of magnetic character of days and monthly averages of ultra-violet radiation, linear scales were found to be better than logarithmic scales.

Correlation by Years

In spite of the irregularities in the long-wave field strength curves in Fig. 2, it is evident that during the sunspot maximum period the radio

* See Appendix I.

field strength tended to be higher than during the years of sunspot minimum. The same applies to solar limb-prominences, magnetic activity, and ultra-violet radiation.† The irregularities in the radio telephone field strength curves are more nearly duplicated in the curves

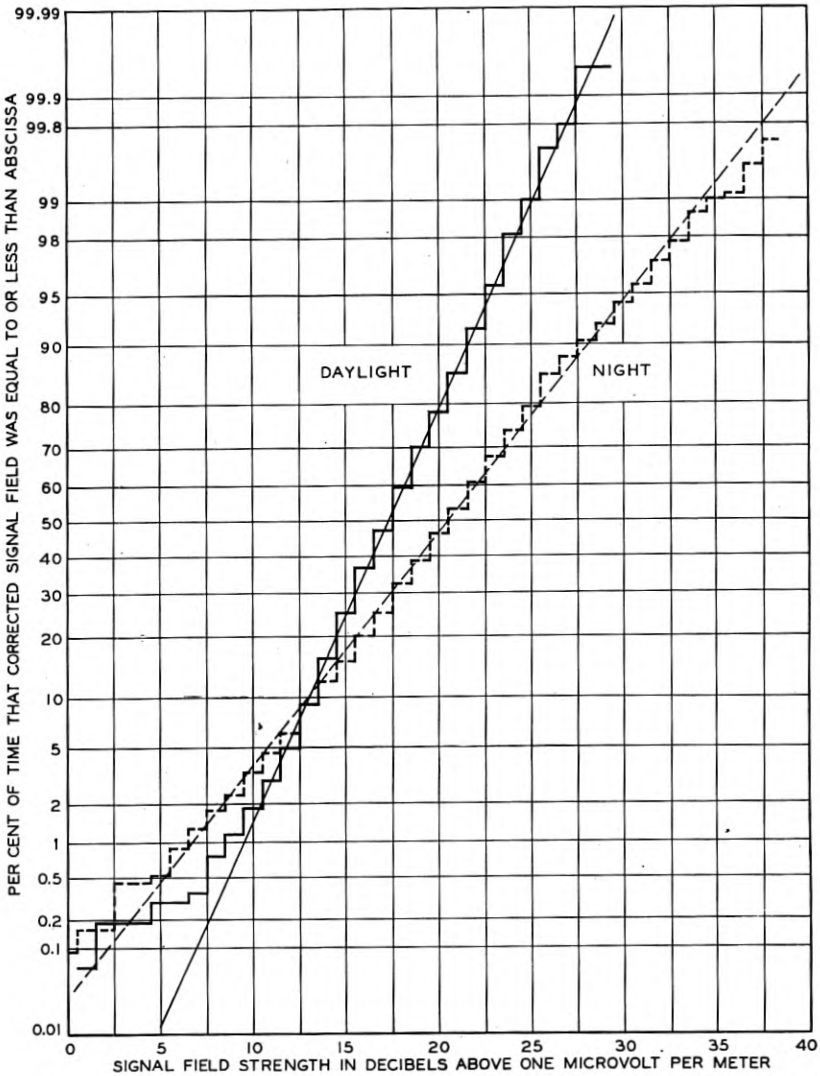


Fig. 1—Cumulative distribution of hourly observations of the field strength of WNL—60 kc. at Cupar, Scotland, in 1933, corrected to 375 amperes. (1699 observations for all-daylight path, 1146 observations for all-night-time path.)

† The measure of ultra-violet radiation is the ratio of intensity of ultra-violet ($\lambda = 0.32\mu$) to green ($\lambda = 0.50\mu$). Ratio for June 1924 = 1.

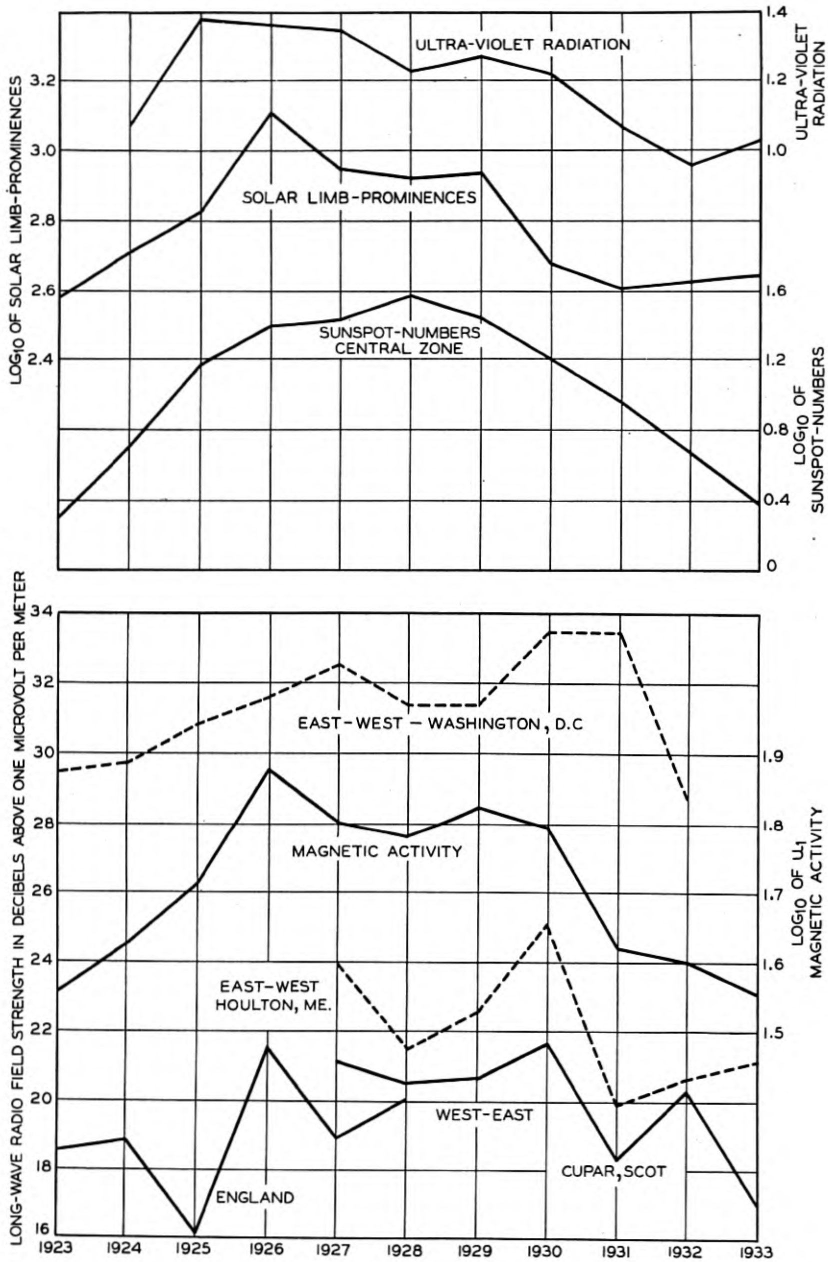


Fig. 2—Annual averages of solar phenomena, terrestrial magnetic activity, and long-wave radio transmission during one 11-year sunspot cycle.

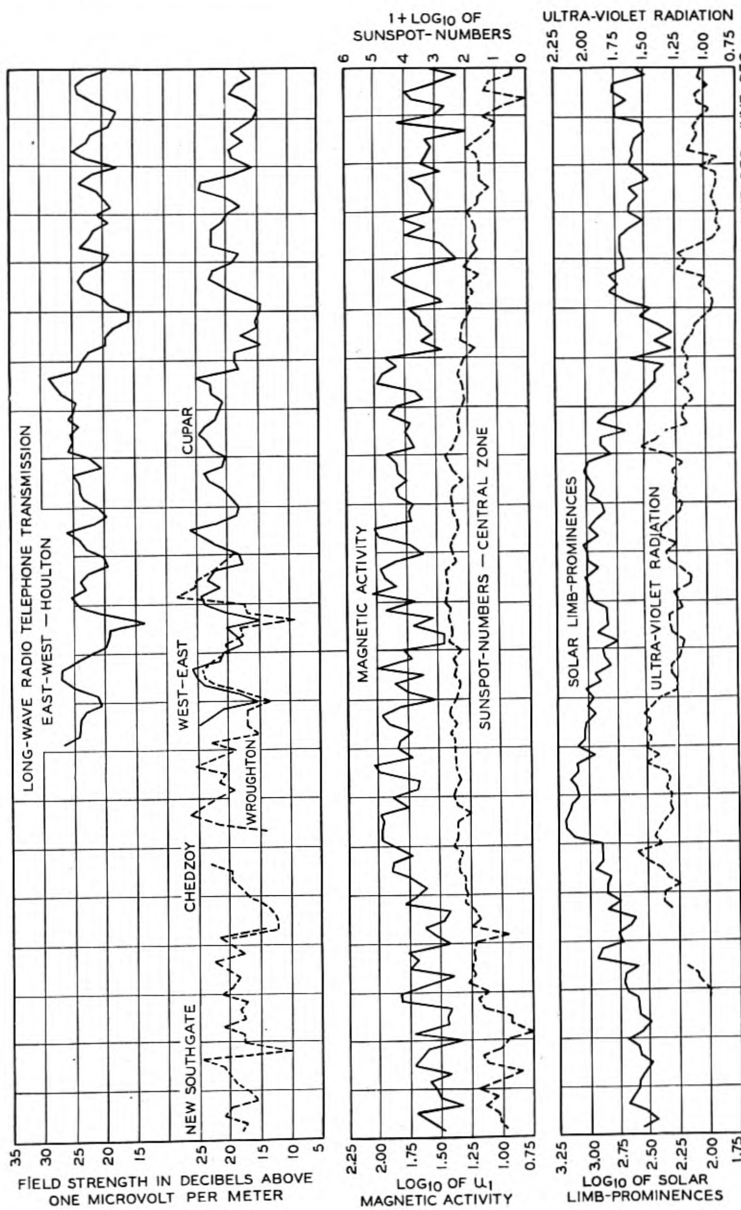


Fig. 3—Monthly averages of solar phenomena, terrestrial magnetic activity, and long-wave radio transmission during one 11-year sunspot cycle.

of magnetic activity and solar limb-prominences than in the sunspot and ultra-violet curves. The curve of reception at Washington, D. C., of signals from several European telegraph stations on longer wavelengths (15 to 30 kc.) followed the Houlton and Cupar curves fairly well except in 1931.

The decrease in long-wave field strength apparently lagged a year or two behind the decrease in sunspot-numbers. This may perhaps be explained by the lower heliographic latitude of the sunspots near the minimum of the solar cycle, offsetting the decrease in number of spots.

Correlation by Months

Monthly averages of the same phenomena as shown in Fig. 2 are plotted in Fig. 3, omitting the curve of radio telegraph reception at Washington. There seems to be little or no obvious correlation between the field strength curves and the other curves except for the one of magnetic activity. Sunspots in the central portion of the solar disc are believed to have the major influence on terrestrial phenomena,[4, 5, 6, 7] and therefore the sunspot-numbers employed in this analysis are those for the central zone * of the sun. Solar limb-prominences would hardly be expected to have considerable influence † on the terrestrial phenomena, although some relation is suggested by Fig. 2. Monthly averages of ultra-violet radiation also offer little explanation of the variations of long-wave fields.

Magnetic activity easily gives the best correlation with long-wave fields on a month-to-month basis. A scatter diagram was constructed in which the West–East and East–West monthly averages of daylight long-wave radio transmission were correlated with terrestrial magnetic activity for the years 1927–1932. For WNL the coefficient of correlation [8] is 0.526 ± 0.059; for GBT and GBY, 0.747 ± 0.067, indicating a high degree of correlation. Such correlation means either that magnetic activity affects radio transmission or that both are affected similarly by a common cause of disturbance.[9]

The coefficient of correlation of magnetic activity with the monthly averages of WNL and 2XS at New Southgate in 1923 and 1924, at Chedzoy in 1925, and at Wroughton in 1926, 1927, and 1928 is 0.38 ± 0.10.

* From 1917 to 1928, inclusive, the "central zone" of the sun was defined as that part of the sun's surface included between two meridians situated 30° on either side of the central meridian. (The central meridian is that meridian of the sun which bisects the sun's disc.) Beginning in the year 1929 the central zone is defined as the area on the sun's disc enclosed by a central circle having a diameter half that of the disc.

† Solar limb-prominences may have a more direct relation to sunspots without having any significant relation to terrestrial phenomena.

Very striking and significant results were obtained when seasonal averages of magnetic activity and radio field strength were plotted in Fig. 4 to indicate seasonal trends. Except for the magnetic sub-maximum in May the curves of field strength and magnetic activity

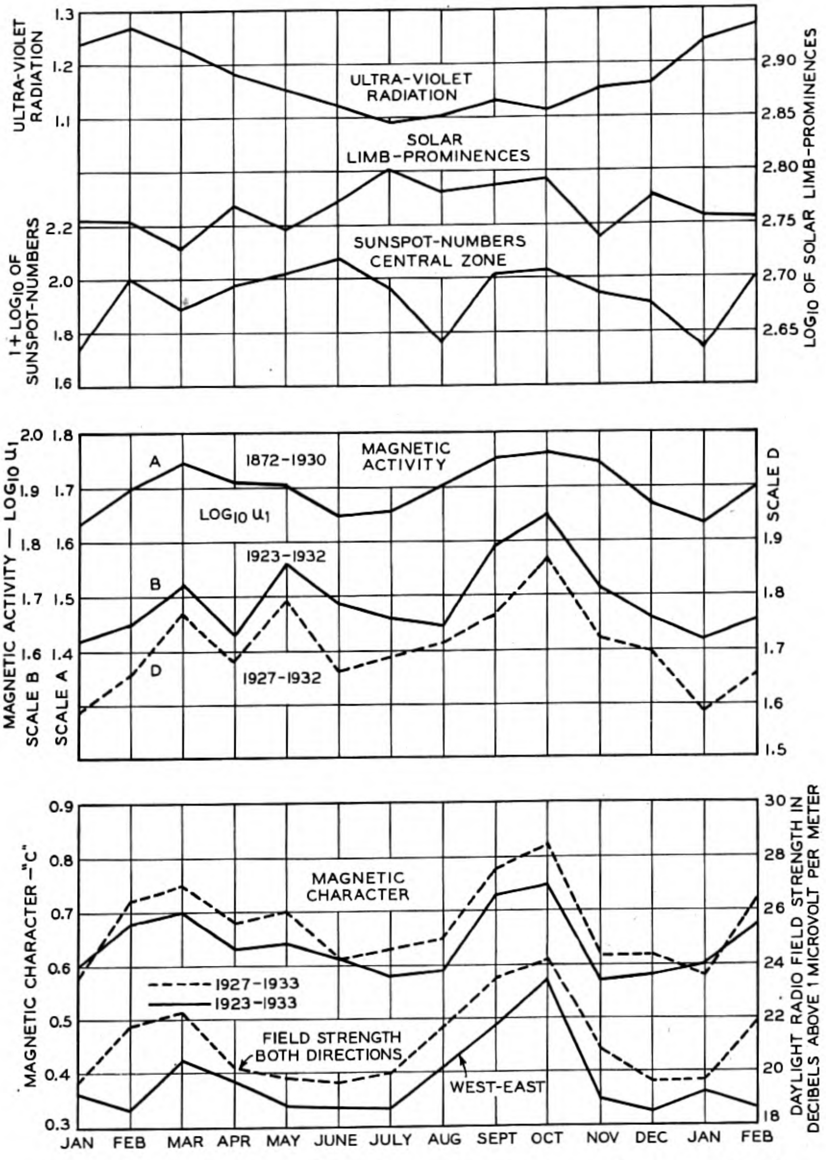


Fig. 4—Seasonal averages of long-wave radio field strength compared with solar phenomena and terrestrial magnetic activity.

are closely in phase. While the two curves of average magnetic characterization ("*C*") are considered not so reliable * for this analysis as the u_1 curves, the same equinoctial maxima appear in all magnetic curves. However, seasonal averages of solar limb-prominences, sunspot-numbers, and ultra-violet radiation apparently have no significance in relation to radio field strength. Bartels [10] found that there is no real seasonal cycle of sunspot-numbers. Cupar and Houlton observations were combined † for the years 1927 to 1933 to make the dashed curve.

West–East average transmission for eleven years is shown by the solid curve at the bottom of Fig. 4. The February dip was recorded at Cupar, Scotland, in 1931 only but was previously observed at the receiving locations in England (also during four of the seven years at Houlton). The $\log_{10} u_1$ curve for 1872–1930 was derived from data published by J. Bartels [10] in a comprehensive analysis of magnetic data, which he critically examined for the reality of seasonal variability. The other $\log_{10} u_1$ magnetic activity curves ‡ cover the first ten of the eleven years for which the radio data are available in one direction, and the first six of the seven years for which radio data are available in both directions, East–West and West–East.

Correlation with Magnetic Storms

The average daylight radio field strength of long-wave signals does not reach a maximum until five to eight days § after the peak of a long magnetic storm, nor does the average night-time field reach a minimum until about the same time. This is shown (Fig. 5) by a study of 25 storms of four or five days duration in the four years from 1930 to 1933, inclusive. The incidence of succeeding storms did not permit the study to include a longer time after the storms had subsided, so that less than one complete cycle of field variation is all that can be shown with any confidence. Data for WNL, GBT, and GBY were combined to obtain the trend curves for the radio field strength. A study of 16 periods (in the years 1927–1933) of one and two months' duration, including one to three significant magnetic storms, showed that the difficulty involved in extending this kind of study to individual storms lies primarily in the lack of highly accurate radio data.

* See Appendix I.

† See Appendix II.

‡ These curves were drawn before the data for magnetic activity (u_1) for 1933 were available. The 1933 data became available shortly before the paper was released for publication but do not materially change the overall picture.

§ C. N. Anderson [11] showed a similar delay in the maximum increase of daylight field following severe magnetic storms in 1927 and 1928. I. J. Wymore [12] found a delay of about two days for signals of somewhat longer wave-length.

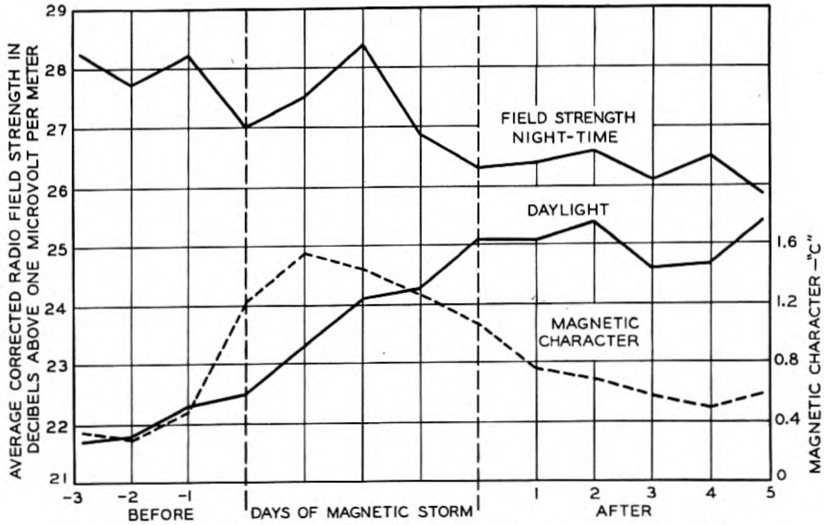


Fig. 5—Average daylight and night-time long-wave radio field strength before and after 25 long magnetic storms.

27-Day Intervals

Sunspot-numbers and terrestrial magnetic disturbances tend to be repeated at 27-day intervals.[10] A recurrence correlation for "radio character" and magnetic character has been published by A. M. Skellett,[13] who based the radio data upon the relative intensity from day to day of the disturbances of the short-wave telephone circuits between New York and London. A recurrence chart for long-wave radio is shown in Fig. 6, based upon the deviation from the 50-day moving average of the average daylight radio field strength of WNL (60 kc.) at Cupar for seven years beginning with March 1927. A measure of the international magnetic character figure ("*C*") for each day in the same period is likewise shown in this figure.

It is evident that the 27-day recurrence phenomenon is obscured on long waves, probably because magnetic storms are followed by prolonged effects on the long-wave fields, as was pointed out in connection with Fig. 5.

Correlation by Days

The average daylight field strength on long waves is improved on days of magnetic disturbance,[14] as compared with calm days.[15] Cumulative distribution curves of daily averages of daylight field strength of WNL (Rocky Point to Cupar, Scotland) for 1930 and 1933 are shown in Fig. 7, for calm and disturbed days.* Similar data for east-to-west

* A magnetically calm day is defined for the purposes of this paper as a day for which the international magnetic character figure "*C*" is from 0.0 to 0.3. A magnetically disturbed day is here characterized by a figure from 1.2 to 2.0.

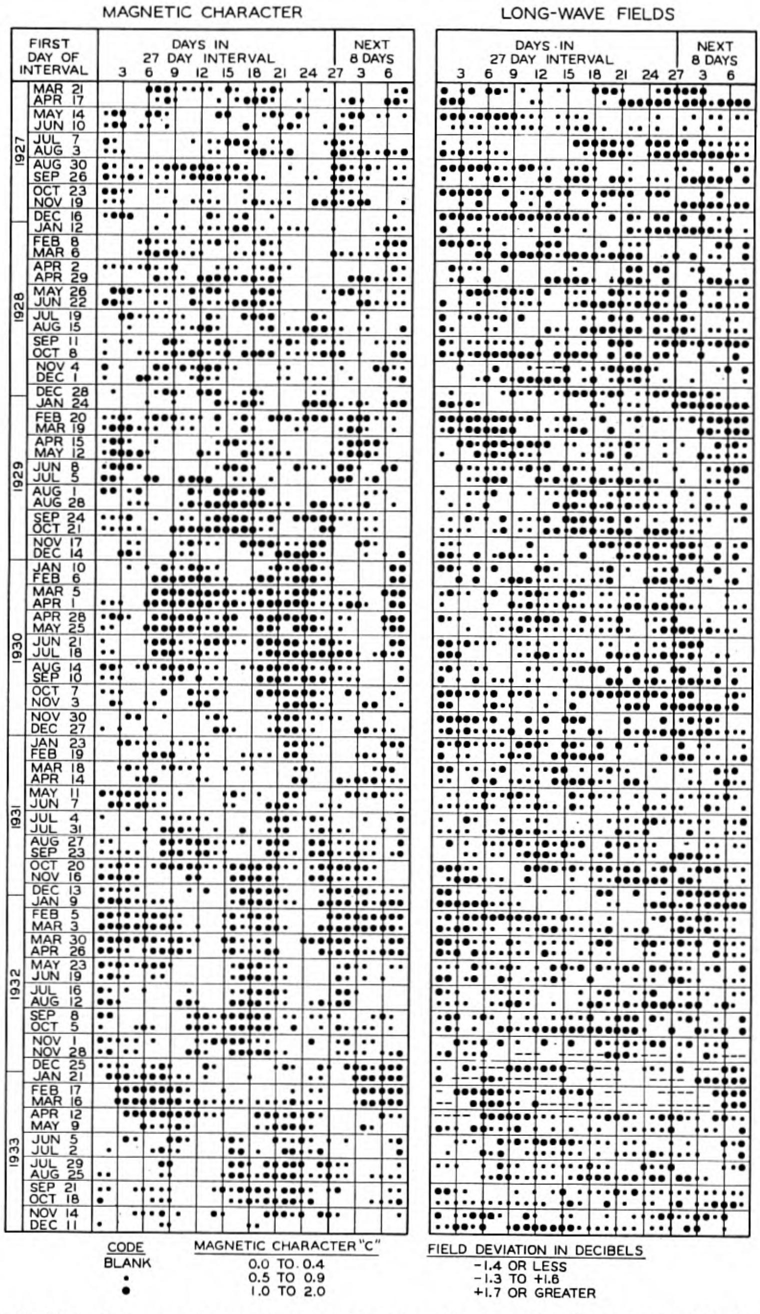


Fig. 6—Twenty-seven-day recurrence diagram of magnetic character and deviation of average daylight field of WNL from 50-day moving average.

transmission from Rugby to Houlton were obtained but are not reproduced. (It will be observed that the antenna current to which the data are corrected is different for different years. This basis results in a convenient spacing of the distribution curves to prevent

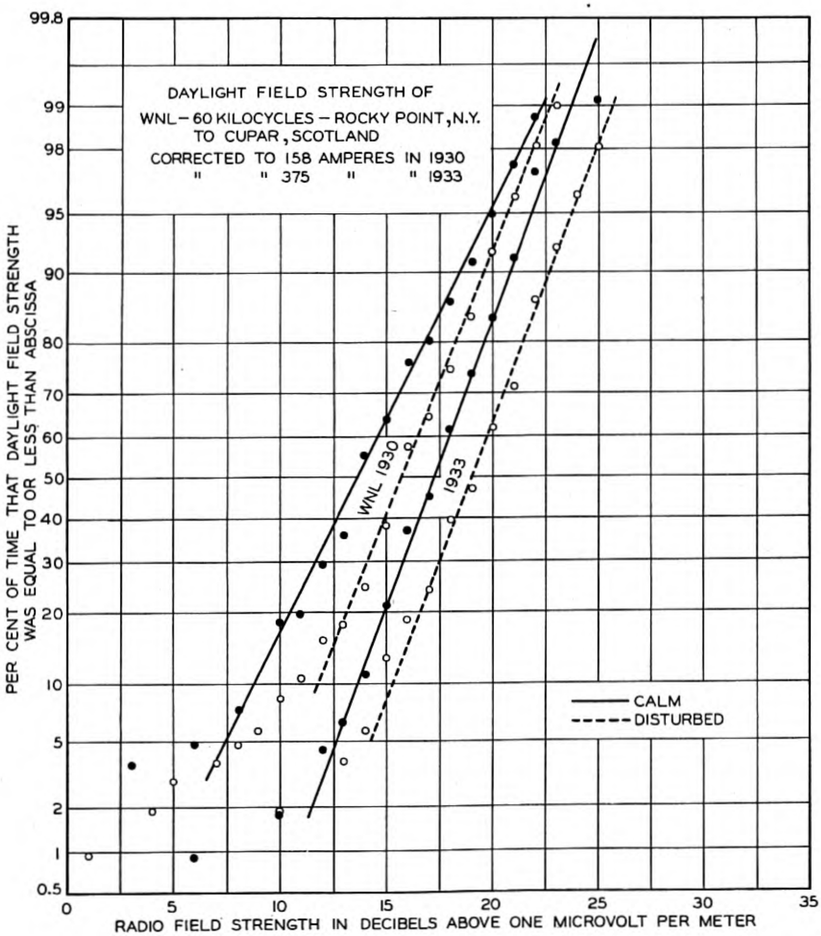


Fig. 7—Cumulative distribution of daily averages of long-wave radio field strength on days of magnetic calm and on days of magnetic disturbance. Daylight transmission path, WNL.

overlapping.) The year 1930 was magnetically more disturbed than 1933. Because of the few days' lag (Fig. 5) of the increase in daylight field strength behind magnetic storms, the full effect that accompanies magnetic disturbances is not revealed by these curves.

No consistent relationship between night-time long-wave fields and magnetic disturbance is obtained from similar analyses. Lowered night-time fields on East–West (GBT, GBY) transmission were found to accompany magnetic disturbances in both quiet and disturbed years. However, from Fig. 8, it is seen that on West–East (WNL) trans-

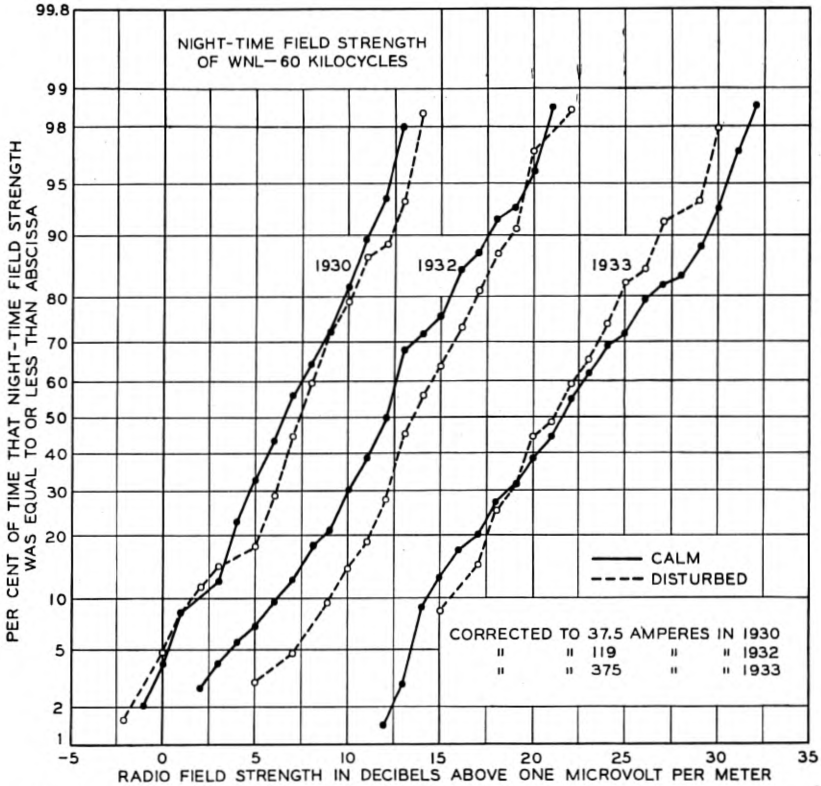


Fig. 8—Cumulative distribution of daily averages of long-wave radio field strength on days of magnetic calm and on days of magnetic disturbance. Night-time transmission path, WNL.

mission, the same result appears only in 1933, whereas in 1930 and 1932 night-time fields averaged higher on disturbed days than on calm days. No reason for this inconsistency has been found.

Correlation with Weather

Other investigators have shown some relation between long-wave field strength and temperature [16] and barometric pressure at short distances.[17] Because of the greater range of weather conditions, the relation is expected to be more obscure on long distance transmission. However, on European signals (15 to 20 kc.) received at Washington from 1924 to 1929, during the periods from September 15 to May 15, there appears [18] "to be a decided drop in signal intensity (measured at 10:00 A.M. and 3:00 P.M.) coincident with heavy rainfall, rising tempertture, and falling barometer, at the receiving station at the time of general storms, and an increased signal intensity as the storm center passes and is followed by clear and colder weather." A conclusion was presented [18] that "either radio waves are subject to large absorption in the lower atmosphere, due to a degree of ionization not known to exist, or surface storms are influenced by conditions existing at great heights."

Disturbed weather conditions may be part of the reason for disturbed transatlantic long-wave (60 kc.) radio reception, but the authors favor the hypothesis that local weather storms are more a symptom of disturbed conditions in higher regions. The same could be said of earth currents.[19] However, since reflection probably occurs at the surface of the earth one or more times on the transatlantic paths, as will be discussed more fully in a paper now in preparation, signals using these paths probably encounter surface weather conditions and earth current influences at least three times.

Correlation with Meteor Showers

An attempt was made to find some correlation between meteor showers and long-wave radio transmission. The sketchy nature of the available meteor data does not provide a satisfactory basis for such a study. Comparison of monthly average meteor hour rates [20] with monthly average radio transmission showed little or no promise of significant relations. Although meteors probably do exert an influence on radio transmission [13, 21, 22] by causing random variations of the received signal, due to changes in ionization of the transmission path, the evidence for major changes in field strength at times of meteoric activity apparently is lacking in our data.

During the four-month period centering on the Perseid shower,* Pickard [23] observed a considerable depression in the day reception of long-wave signals from Nauen (23 kc.) and a considerable increase in medium-wave signals at night from WBBM (1330 kc.), with maximum effects close to the day of meteor shower maximum. No such effect was detected, however, in the data on the long-wave transatlantic radio telephone circuits for the period from 1923 to 1933. In fact, as shown in Fig. 4, if the Perseid shower could be credited with having con-

* According to Olivier in Bulletin No. 8, the American Meteor Society (quoted also in reference 23), the Perseid shower is by far the longest and largest annual meteor shower and extends about 17 days on each side of August 11.

siderable effect on long-wave radio transmission, the effect would apparently be an increase in field strength in daytime, opposite to the result observed by Pickard on signals from Nauen. For the period covered by the curves in Fig. 4, the increased daylight field strength in August is not explained by correlation with the curves of magnetic activity.

Nature and Sources of Data

The long-wave radio telephone data used in this study are measurements of field strength at the receiving stations, and each individual observation is a simple physical measurement. The data taken before 1927, although relatively meager, were observations made under controlled conditions, a single-frequency signal being employed. Since 1927 the data, while much more abundant, had to be taken under operating conditions with speech transmission. However, since such measurements can be made on a commercial circuit only when in the judgment of the operating forces it is desirable and feasible to establish that circuit, it is inevitable that the overall picture will be distorted by the requirements of practical circuit operation. This could readily obscure such minor effects as correlation with meteoric showers, and would be likely to introduce some discrepancies between day and night behavior.

Sources of the data used in Figs. 2 and 3 are as follows:

Ultra-violet and sunspot data from *Terrestrial Magnetism*, **39**, 234, September 1934.
Solar limb-prominences data from *Astronomische Mitteilungen*, **130**, 217, 1933, and from **131**, 23, 1934.
Data for the Washington measurements—for 1923, figure by correspondence with the Bureau of Standards; for 1924 to 1930, reference 24; for 1931–1932, reference 25; observations were discontinued in 1933.
Magnetic character data, "*C*" measure, from volumes 29 to 39, inclusive, of *Terrestrial Magnetism*.
Magnetic activity, u_1 measure, from references 10, 26, and 28.

Acknowledgments

The authors wish to thank the personnel of the American Telephone and Telegraph Company and British General Post Office who made the observations and supplied the radio transmission data used in this paper. Acknowledgments are also due the Carnegie Institution of Washington, the United States Bureau of Standards, and Dr. Charles P. Olivier for other information used by the authors in correlating and interpreting the results.

Bibliography

1. H. Plendl, "Concerning the Influence of the Eleven-Year Solar Activity Period Upon the Propagation of Waves in Wireless Telegraphy," *Proc. I. R. E.*, **20**, 520, March 1932, and references.

2. G. W. Kenrick and G. W. Pickard, "Summary of Progress in the Study of Radio Wave Propagation Phenomena," *Proc. I. R. E.*, **18**, 649, April 1930, and references.
3. Lloyd Espenschied, C. N. Anderson, and Austin Bailey, "Transatlantic Radio Telephone Transmission," *Proc. I. R. E.*, **14**, 7, February 1926, *B. S. T. J.*, July 1925.
4. C. N. Anderson, "Correlation of Long-Wave Transatlantic Radio Transmission with Other Factors Affected by Solar Activity," *Proc. I. R. E.*, **16**, 297, March 1928.
5. L. W. Austin, "Solar Activity and Radio Telegraphy," *Proc. I. R. E.*, **20**, 280, February 1932, and references.
6. G. W. Pickard, "The Correlation of Radio Reception with Solar Activity and Terrestrial Magnetism," *Proc. I. R. E.*, **15**, 83, February 1927; also, **15**, 749, September 1927.
7. G. W. Pickard, "The Relation of Radio Reception to Sunspot Position and Area," *Proc. I. R. E.*, **15**, 1004, December 1927.
8. H. L. Reitz, "Mathematical Statistics" (book), Open Court Publishing Company, Chicago, 1927, Chapter IV.
9. H. B. Maris and E. O. Hulburt, "A Theory of Auroras and Magnetic Storms," *Phys. Rev.*, March 1929.
10. J. Bartels, "Terrestrial-Magnetic Activity and Its Relations to Solar Phenomena," *Terr. Mag.*, **37**, 1, March 1932.
11. C. N. Anderson, "Notes on the Effect of Solar Disturbances on Transatlantic Radio Transmission," *Proc. I. R. E.*, **17**, 1528, September 1929.
12. I. J. Wymore, "The Relation of Radio Propagation to Disturbances in Terrestrial Magnetism," *Proc. I. R. E.*, **17**, 1206, July 1929.
13. A. M. Skellett, "Disturbances in Radio Transmission," *Bell Laboratories Record*, **13**, 164, February 1935.
14. H. B. Maris and E. O. Hulburt, "Wireless Telegraphy and Magnetic Storms," *Proc. I. R. E.*, **17**, 494, March 1929.
15. C. N. Anderson, "Notes on Radio Transmission," *Proc. I. R. E.*, **19**, 1150, July 1931.
16. L. W. Austin and I. J. Wymore, "Radio Signal Strength and Temperature," *Proc. I. R. E.*, **14**, 781, December 1926.
17. K. Sreenivasan, "On the Relation Between Long-Wave Reception and Certain Terrestrial and Solar Phenomena," *Proc. I. R. E.*, **17**, 1793, October 1929.
18. I. J. Wymore-Shiel, "A Correlation of Long-Wave Radio Field Intensity with the Passage of Storms," *Proc. I. R. E.*, **19**, 1675, September 1931.
19. Isabel S. Bemis, "Some Observations of the Behavior of Earth Currents and Their Correlation with Magnetic Disturbances and Radio Transmission," *Proc. I. R. E.*, **19**, 1931, November 1931.
20. Charles P. Olivier, "Meteors" (book), Williams and Wilkins, Baltimore, 1925, Chapter 16, p. 182.
21. A. M. Skellett, "The Ionizing Effect of Meteors," *Proc. I. R. E.*, **23**, 132, February 1935.
22. A. M. Skellett, "The Ionizing Effect of Meteors in Relation to Radio Propagation," *Proc. I. R. E.*, **20**, 1933, December 1932.
23. G. W. Pickard, "A Note on the Relation of Meteor Showers and Radio Reception," *Proc. I. R. E.*, **19**, 1166, July 1931.
24. L. W. Austin, "Tables of North Atlantic Radio Transmission Conditions for Long-Wave Daylight Signals for the Years 1922–1930," *Proc. I. R. E.*, **20**, 689, April 1932.
25. L. J. Briggs, "North Atlantic Radio Transmission Conditions for Long-Wave Daylight Signals for the Years 1931 and 1932" (Letter to the Editor), *Terr. Mag.*, **38**, 67, March 1933.
26. J. Bartels, "Terrestrial-Magnetic Activity in the Years 1931 and 1932," *Terr. Mag.*, **39**, 1, March 1934.
27. Charles Chree, "The Criterion of Magnetic Disturbance," *Terr. Mag.*, **28**, 33, March 1923.
28. J. Bartels, "Terrestrial-Magnetic Activity in the Year 1933 and at Huancayo," *Terr. Mag.*, **40**, September 1935.

APPENDIX I

Measures of Terrestrial Magnetic Activity "C" and "u_1"

Characterization (C)

"Terrestrial-magnetic activity at a given station, and in a certain interval, may be defined as an expression for the frequency and intensity of magnetic disturbances in that interval. There are many ways in which this general definition may be expressed as a numerical measure. *Characterization*, the simplest, is now widely used. In this measure every observatory assigns, from the character of its photographic records, to each interval of 24 hours, between successive Greenwich midnights, a character-figure, '0' for quiet, '1' for moderately disturbed, and '2' for greatly disturbed days. The average for all collaborating observatories (the number of which increased from 30 to about 45, since this measure was begun in 1906) is the international magnetic character-figure C." [10]

"A primary desideratum is to arrive at a clear idea of exactly what it is we want to measure. If our object is simply to discriminate between the days of a single month, with a view to selecting for special purposes the five quietest or five most disturbed days of the month, it is very doubtful whether the existing scheme of international 'character' figures can be improved on. Its simplicity and the small amount of labor it entails are great recommendations. The disadvantages it seems to me to possess are: 1. The significance of any particular 'character' figure, *e.g.*, 1.5, is variable; it connotes decidedly less disturbance in a quiet than in a disturbed year. Also, while the mean 'character' figure for the year does to a certain extent wax and wane with disturbance, the variation seems to me inadequate. 2. The assigning of 'character figures' at an individual station is largely a psychological process, depending on the temperament and knowledge of the judge. The standards in use at different stations at the same time are widely different, and the standard in use at any particular station may vary largely from time to time. The 'character' figures supplied by any two stations do not suffice for a satisfactory intercomparison of the stations, and if we wish to compare one year or season with another the international 'character' figures leave a great deal to be desired." [27]

The "C" measure of magnetic activity may be inferior as a basis for comparing years, but is relatively safe for comparing short intervals such as days within a month. For comparing longer intervals a different measure is desirable, having a homogeneous standard of evaluating magnetic activity from year to year and for all time. This standard measure was called "u," from which was derived the "u_1" measure used in this paper.

u_1 *Measure*

"A short *definition of u* may be repeated as the average change, taken without regard to sign, measured in the unit 0.0001 c.g.s. = 10γ, of the daily means of horizontal intensity on the magnetic equator of the earth. The monthly means of u_1 are derived from those of u:

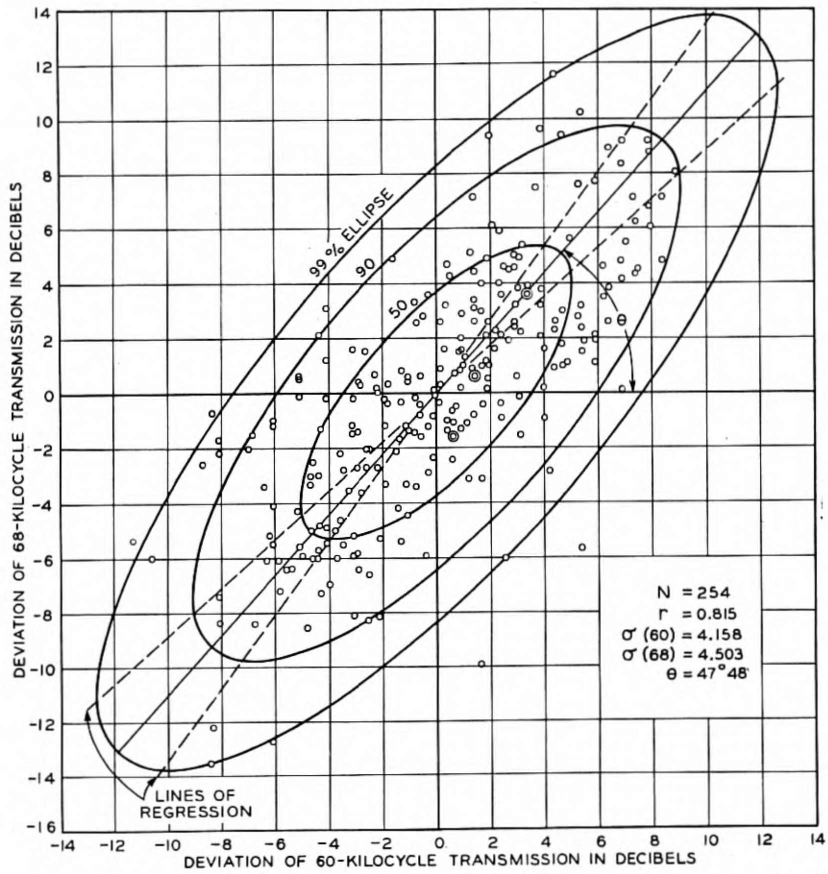


Fig. 9—Correlation diagram of scattering from mean of daily averages of radio transmission over all-daylight paths for the 254 days from May 5, 1931 to March 31, 1932, when GKA–GBY (Rugby, England) was transmitting on 68 kc. and WNL (Rocky Point, N. Y.) was transmitting on 60 kc.

For values of u from 0.0 to 0.6, $u_1 = 100u - 30$.

For higher values,[10]

$$0.6 \text{ to } 1.6,\ u_1 = 30 + 100(u - 0.6) - 30(u - 0.6)^2,$$
$$1.6 \text{ to } 3.6,\ u_1 = 100 + 40(u - 1.6) - 10(u - 1.6)^2,$$
$$3.6 \text{ to } \infty,\ u_1 = 140,$$

u_1 increases less rapidly, approaching asymptotically the limiting value 140. The quantity u_1 has been introduced in order to obtain a measure of activity which has a frequency-distribution similar to that of the relative sunspot-numbers, and therefore is more suitable for research on correlations between terrestrial-magnetic activity and solar activity." [26]

APPENDIX II

Data for East–West transmission were combined with data for West–East transmission for two illustrations, Figs. 4 and 5, given in this paper. Two unpublished studies by the authors of this paper showed that the correlation between transmission from Rugby to Houlton on 68 kilocycles and transmission from Rocky Point to Cupar on 60 kilocycles was very high, 0.80 ± 0.01. The scatter diagrams for these studies were so similar that only one is shown here, Fig. 9. For transmission on 60 kilocycles in both directions the correlation was 0.76 ± 0.02. These results are considered sufficient justification for combining East–West and West–East radio transmission data to obtain average curves.

TECHNICAL DIGESTS

With this issue the first of a series of briefer articles, known as "Technical Digests," is being introduced. These will review papers by Bell System authors which appear in the current numbers of other scientific or engineering periodicals.

The purpose of these digests, which in length will run to several times that of the usual abstract, may be given as twofold. By eliminating the need hitherto felt that numerous articles be reprinted in full, a larger variety of subjects will find compass in a given number of Technical Journal pages. Secondly and more important, it is hoped that the digests will succeed in so outlining the important features of their respective articles as to make them available to a much wider circle of readers than would have opportunity for a detailed reading or study of the original.

Superiorities of Lead-Calcium Alloys for Storage Battery Construction *

RECENT investigations, conducted at the Bell Telephone Laboratories and elsewhere, have demonstrated that the lead-antimony alloys almost universally employed in storage cell construction are far from ideal for the purpose from the electrochemical standpoint. It has been shown that *in the course of normal operation* of the present type cell, antimony is leached out of the positive electrode, passes through the solution and deposits on the negative, where it promotes "local action" and self-discharge. Also, it has been demonstrated that stibine is generated in perceptible amounts by the present type battery on over-charge.

The continued use of lead-antimony alloys for over fifty years has been due primarily to their desirable metallurgical and physical characteristics and the fact that other equally satisfactory alloys of lead have not been available. Electrochemical theory indicates that for use in storage cell construction, lead should be alloyed only with metals *electronegative* to it. The alloying constituents should have little tendency to diffuse or segregate at the normal operating temperatures reached in a cell. The resulting alloy should be considerably stronger than lead, easily cast, and resistant to electrolytic corrosion. It should also have high electrical conductivity and small solidification and thermal contraction.

As a result of a comprehensive investigation of lead-calcium alloys in connection with cable sheathing materials, data were accumulated at the Bell Telephone Laboratories which suggested the use of certain of these alloys for storage battery grids and plates. Tests have been conducted to determine the value of this suggestion, and with very promising results.

In the course of the cable sheath studies, the thermal equilibrium diagram of lead-calcium alloys containing a very small percentage of the calcium component was determined and is illustrated in Fig. 1. This diagram shows that the amount of calcium soluble in solid lead

* Digest of Two Papers: "The Electrochemical Behavior of Lead, Lead-antimony and Lead-calcium Alloys in Storage Cells" by H. E. Haring and U. B. Thomas, and "Some Physical and Metallurgical Properties of Lead-calcium Alloys for Storage Cell Grids and Plates" by E. E. Schumacher and G. S. Phipps. These papers are, being presented before the Convention of the Electrochemical Society in Washington, October, 1935, and published in the *Transactions* of the Society.

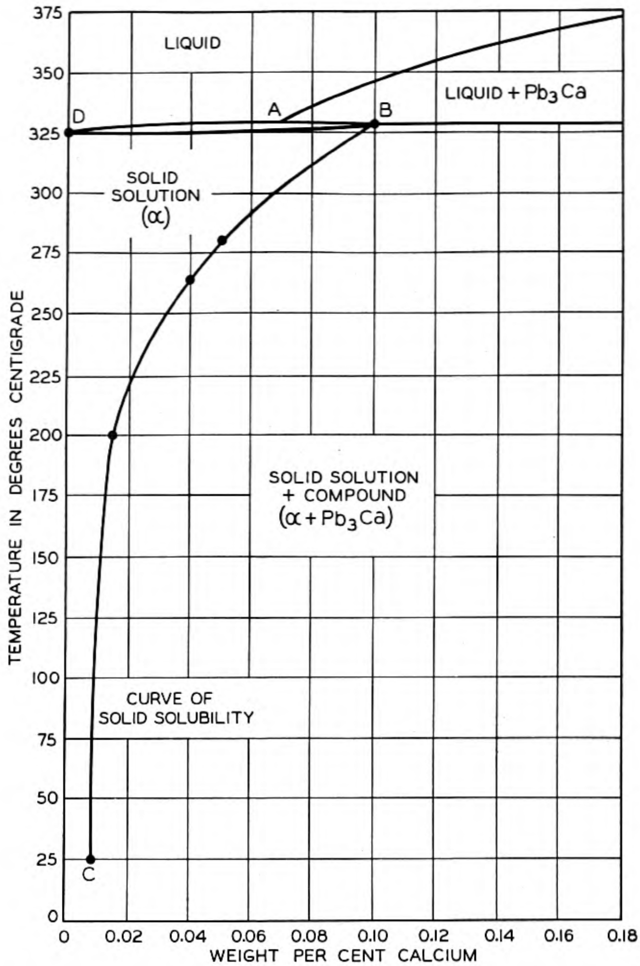


Fig. 1—A partial equilibrium diagram of the system lead-calcium, according to Schumacher and Bouton.

varies continuously from 0.10 per cent at 328° C. to about 0.01 per cent at room temperature. Moreover, when the 0.10 per cent alloy is cooled below 328° C., the calcium precipitates out of the solid solution in the form of Pb_3Ca. This precipitation, while a gradual process, can be hastened by suitable heat treatment. Physically, the effect of the Pb_3Ca molecules highly dispersed throughout the body of the alloy is to strengthen and harden it very materially. Thus, the alloy with a calcium content of 0.10 per cent has a tensile strength of around 8,000 pounds per square inch, a value comparable to that reached by the lead-antimony alloys most generally used in storage battery construction. It has been found, furthermore, that at ordinary temperatures no decrease of tensile strength occurs with age. This maintenance of strength is probably related to the low rate of diffusion of the hardening constituent Pb_3Ca which, being a large and stable molecule, does not diffuse readily in the matrix. The electrical conductivity of this lead-calcium alloy is approximately 20 per cent greater than that of lead-9 per cent antimony, a factor of importance in securing uniform distribution of current when large currents are drawn from a battery.

Experimental cells of both Planté and Fauré (pasted) types have been constructed and are being subjected to a variety of tests. To date, forty-two cells of the starting and lighting type have been pasted, assembled, formed and cycled. This investigation has progressed sufficiently to make it quite evident that the behavior of a complete cell can be predicted with a high degree of accuracy if the electrochemical properties of the structural alloy are known. It has been definitely established that when the negative electrodes of lead-calcium cells of the starting and lighting type (previously subjected to 100 charge-discharge cycles) are charged and allowed to stand one month, they lose only 4 to 5 per cent of their charge as compared to 20 to 25 per cent for lead-antimony cells similarly treated. The efficiencies of lead-calcium cells have been found to be superior to those of lead-antimony cells. No undue corrosion of lead-calcium grids has been observed.

Marine Radio Telephone Service for Boston Harbor *

By F. A. GIFFORD and R. B. MEADER

THERE has been a constantly increasing interest in an inexpensive service for small harbor and coastal craft such as tugboats, private yachts, coastal passenger ships, merchant craft and fishing vessels. This interest became particularly evident in New England in 1931 and since equipment suitable for the purpose had recently been developed by Bell Telephone Laboratories, the New England Telephone and Telegraph Company undertook the establishment of a marine radio telephone service.

A survey consisting of a comprehensive series of field strength measurements on shipboard and at various points along the coasts of Massachusetts and Cape Cod Bays resulted in selecting Green Harbor as the location for a shore station. Green Harbor is in the town of Marshfield, Massachusetts, about 28 miles southeast of Boston.

A commercial survey indicated that initially the service would be of interest chiefly to the Boston fishing industry. Consequently, boat radio telephone equipment was installed on the trawler "Flow" of the Bay State Fishing Company and the service was opened in June, 1932, on a demonstration basis. As tests with this vessel progressed, it became evident that the radio telephone service would fulfill the communications requirements of the fishing industry. It also became apparent that a complete service of this type should include some means for determining the vessel's position at any time by means of radio. Therefore, the development of such equipment was initiated by Bell Telephone Laboratories as an adjunct to the radio telephone service, and the outcome of tests of an experimental model indicates that the problem of providing suitable radio compass equipment in the price range satisfactory to the fishing fleet owners has been satisfactorily solved.

The radio transmitter is a 400-watt crystal-controlled type similar to those designed for use at aviation ground stations and adjusted to operate at a frequency of 2506 kilocycles. This frequency is maintained within limits of better than 0.025 per cent.

In order to combine the two unidirectional radio channels into a two-way circuit suitable for connection to the ordinary wire circuits in the

* Digest of a paper to be published in full in *Communication and Broadcast Engineering*, October, 1935.

land telephone network, apparatus similar to that provided at the terminals of the transatlantic and high seas ship-to-shore radio telephone circuits has been provided. This apparatus includes controls for adjusting the volume of speech into the transmitter from the receiver to the wire lines and the usual voice operated devices (termed "vodas") provided for the suppression of echoes and singing.

The apparatus mentioned, together with a volume indicator, means for talking, monitoring and signaling, and testing apparatus, is as-

Fig. 1—Installation of control unit on fishing trawler.

sembled on one floor-mounted apparatus bay. This, mounted adjacent to the bay containing the receiver and a noise suppression device, constitutes the operating position. This position is continuously attended by a technical operator who adjusts the controls during the progress of each call, guided by indications of the meters provided, to insure the best possible connection under the conditions obtained at the time. Power for the terminal apparatus is supplied by a motor generator set operating from a 110-volt 60-cycle source.

The noise suppression device, termed a "codan" (carrier operated device anti-noise) is employed in connection with the receiver which introduces a predetermined high loss into the audio-frequency portion of the radio receiver during intervals when there is no incoming carrier, and insures relatively quiet conditions on the receiving line. When a carrier is received, the codan action removes this loss, allowing the speech with which the carrier is modulated to pass from the receiver

Fig. 2—Marine radio compass installed on a trawler.

output to the receiving line. This device makes it possible to deliver higher speech volumes to the telephones on shore since it prevents the operation of the receiving vodas relays on radio noise during the idle intervals. It also prevents the high radio noise which would otherwise result due to the increase in gain inserted by the automatic volume control whenever the carrier is interrupted.

It can be seen from this discussion that the use of the codan presupposes suppression of the carrier of the distant transmitter except when the user wishes to talk. This method of operation has been adopted for this type of marine radio telephone service.

A 10-kw. 220-volt three-phase 60-cycle alternator driven by a Buffalo gasoline engine has been provided to furnish the necessary emergency power supply in case the normal commercial supply fails.

The frequency designated by the Federal Communications Commission for use by ships communicating with the shore through the Green Harbor radio telephone station is 2110 kilocycles. This carrier frequency is maintained within limits of 0.025 per cent.

Two crystals are provided and the circuit arranged so that the receiver may be quickly adjusted to operate on either of two frequencies by means of a local mechanical or an electrically operated remote control. The receiver is so designed as to make possible boat-to-boat conversations on a separate frequency.

The signaling unit which is normally connected to the output of the radio receiver consists of a selector operated under the control of an arrangement of relays which in turn are controlled by incoming signal pulses of 600 and 1500-cycle tones. The bell on each vessel is operated only in response to the particular code of pulses to which the selector is adjusted. Arrangements are also included so that the vessels of any one fleet may be called simultaneously.

A motor generator set operates continuously while the vessel is standing by for the reception of signals and furnishes 12- and 200-volt power for the operation of the radio receiver and signaling unit. A second motor generator set is automatically started when the handset is lifted from the switch hook to place a call or in response to an incoming signal, and furnishes power to operate the transmitter. On several of the smaller boats having 32-volt power supply with wide voltage fluctuations, power supply equipment consisting of two dynamotors operated from a 12-volt battery charged from the vessel's storage battery has been employed successfully.

The control unit for the radio telephone consists of a small panel on which are mounted a switch for turning the set on and off, a meter for indicating antenna current, a manual volume control, pilot lamp signals and a bell for announcing incoming calls. A special handset with push button completes the control unit assembly.

At Boston, the marine radio telephone traffic is handled at two positions on the outgoing toll board especially modified for this purpose. The wire lines from the Green Harbor station terminate at this point, and calls from vessels can be switched by the operator to any point connected to Bell System facilities. The normal wire lines are three loaded cable pairs. One pair is used for transmission from the shore telephone to a boat, the second conducts speech received from a boat through the toll position to the land line telephone, and the third is employed as an order wire for communication between the operator at the marine position and the technical operator at the Green Harbor station. All of these circuits are duplicated over an alternate route for use in case of trouble on the normal facilities.

The marine operator dials the code assigned to the vessel desired. The dialing operation produces the desired grouping of 600 and 1500 cycle pulses which modulate the radio transmitter carrier frequency of 2506 kilocycles. The signaling unit on the vessel is actuated by these pulses and the bell rings. The captain raises the handset from the switch hook on the control unit, presses the push button in the handle of the handset and announces the name of his vessel. The operator then completes the connection and the conversation takes place.

In placing a call from boat to shore the captain or member of the crew raises the handset from the switch hook and, after listening to ascertain that no conversations are in progress, presses the push button and calls "marine operator." The marine operator who is normally monitoring on the channel ascertains the name of the calling vessel, the shore station desired and other necessary details, and while the calling party holds the line proceeds to call the land line telephone and establish the connection.

When a person on one boat wishes to talk with a person on another boat, the procedure in placing the call and establishing the connection is the same as in the case of a ship-to-shore call, except that when both are prepared to talk, the technical operator at Green Harbor operates a by-pass key which connects the radio receiver output and radio transmitter input without including the voice operated device and other equipment associated with the land circuits. The land line is bridged onto the circuit so that the marine operator may be advised of any difficulties which arise in carrying on the conversation.

During the more than two years that the system has been in experimental service, the transmission results up to distances of 500 miles from the shore station have been quite satisfactory. Of course, during periods of abnormally heavy static the normal range is somewhat reduced. The service is available at all times, but practically all business is handled between the hours of 8 A.M. and 6 P.M. so that the relatively poor atmospheric conditions usually existing during summer nights do not adversely affect the radio telephone traffic. However, experience has indicated that calls originated during such periods from vessels within the normal range can be handled satisfactorily. During some periods of favorable atmospheric conditions experimental transmissions over distances greatly in excess of the normal range have been successfully conducted.

On fishing vessels the radio telephone equipment is accessible for maintenance work only at the conclusion of trips which are usually of about ten days' duration. It is obvious, therefore, that the equipment must be designed for reliable operation over long periods and experience indicates that these requirements have been well satisfied.

Fishing craft normally make use of the service for reporting the details of the catch, for making arrangements to return to port and for talking with other fishing vessels to locate points where fishing is best. The radio telephone has proved of vital importance on several occasions where engine breakdowns necessitated advice from shore in order to make repairs and having replacement parts available upon the vessel's arrival at port. In several instances of sickness and accidents to members of a crew, medical advice has been obtained or the Coast Guard summoned to remove the injured man for quick transportation to a hospital. In one case of severe damage to a trawler as a result of a collision, the Coast Guard were summoned and the owners were able to keep in constant touch with the situation.

Harbor Craft Ship-to-Shore Radio Telephone Service in Puget Sound Area *

By E. B. HANSEN

PUGET SOUND with an area of about 6,000 square miles serves a large region whose principal resources are lumber and fishing. Movement of the raw products to points where the manufacturing and processing are carried on comprises a major part of the short-haul shipping in the Puget Sound area. Large lumber and pulp mills have been permanently established at locations where deep sea shipping and rail facilities are readily available. Logs in the form of large booms are towed to the mills from the sources of supply which extend throughout the area drained by the Puget Sound. The average haul of each boom is about 75 miles and consumes several days. Because of the relatively long time a tug is isolated in each operation, some economical means of dispatching and directing the activities of the tow boats from the land headquarters is desirable. In the case of fishing, there is an even greater need than in the case of log towing for close contact between fishing boat and cannery in order that both the fishing and cannery activities may be coordinated so as to prevent waste of fish during large runs.

From the above, it is clear that some means of communication between harbor craft and land stations would be useful. A preliminary view of the problem indicated that low-powered radio telephone channels from each of these boats to a single land station and thence by wire lines to any telephone, would be the most economical and practical means from the standpoint of the ship owners. Furthermore, the government regulations covering the issuance of operators' licenses for radio telephone transmitters of 50 watts or less power are such that the average member of a ship's crew can qualify after a few hours' instruction, thus making it unnecessary for vessels employing this equipment to carry a special radio operator.

After it was decided that this service was feasible and should be established the initial order of procedure was the selection of the site for the land station.

Due to the fact that it is desirable to locate the transmitter and receiver in juxtaposition, a wide variety of factors had to receive con-

* Digest of a paper published in *Electrical Engineering*, August, 1935, pp. 828–831.

sideration. As to transmission, the distances involved in the harbor area are such as to make it advantageous, from the standpoint of quality and freedom from fading, to make use of a frequency which will permit covering the entire area with the direct or "ground" wave. The signal strengths obtained with ground wave transmission, however,

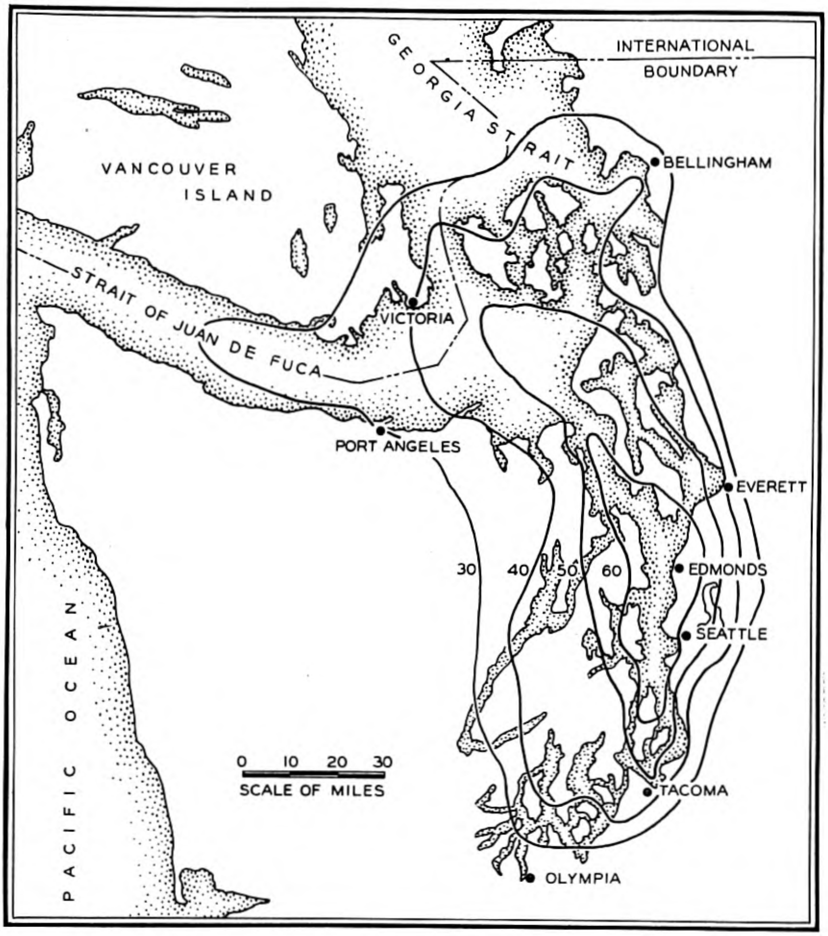


Fig. 1—Field strengths are indicated in decibels above one microvolt per meter.

are affected considerably by the terrain, and the location of the station as much as a mile inland might reduce the over-water range as much as 50 miles. The selection of a receiving site involved a consideration of interference from a variety of sources including high voltage power lines, automobiles and motor boats as well as atmospheric disturbances.

After a survey involving transmitted field strengths and noise measurements, a 13-acre plot of tide and water land having ready access to power supply and telephone connections was selected on Point Edwards about one mile south of Edmonds, Washington, and fifteen miles from Seattle.

Due to the fact that the site for the transmitting and receiving station is under water at high tide, the building housing the radio equipment is of frame construction erected on treated piling.

The transmitting and receiving antennas, located 200 feet apart and near the building are also supported by piling. Both antennas are of the simple vertical type, the former 80 feet in height and supported by a single pole 110 feet high, and the latter 40 feet in height. A tuning unit is mounted in a weather-proofed box on the receiving antenna pole. The transmission line from the receiving antenna to the receiver[1] in the radio equipment building is a special armored and lead covered concentric conductor type.

The equipment at the land station consists of a 400-watt transmitter with a rectifier unit for power supply similar to the type generally used for aviation service. The frequency stability is obtained by the use of a quartz crystal frequency control and will maintain its frequency to better than 0.025 per cent. It is designed for substantially complete modulation of the carrier and under this condition little distortion occurs to the speech frequencies. A tuning unit for tuning the transmitting antenna to resonance is also housed in the building.

As shown on the block diagram Fig. 2, a four-wire circuit is used to connect the radio station and the Seattle office where the circuit becomes two-wire for switching to land subscribers. The equipment includes means for regulating speech volumes; outgoing, so as to always properly load the radio transmitter, and incoming so as to give the shore subscriber the best received volume. A voice-operated device known as the "vodas" provides means for suppressing echoes and singing. This device is essentially the same in principle and in its operation as those used in intercontinental service which have been described in detail in previous publications. In addition to the above units apparatus for monitoring and testing is provided.

The harbor station, moreover, is arranged for remote control operation, equipment being provided so that the transmitter is automatically turned on when the operator inserts a plug in the "Harbor Circuit" jack.

During operation the transmitter is continuously monitored by means of an auxiliary radio receiver in Seattle tuned to the transmitter frequency. In addition, the two-way portion of the voice-frequency

[1] For a description of the receiver see the following digest.

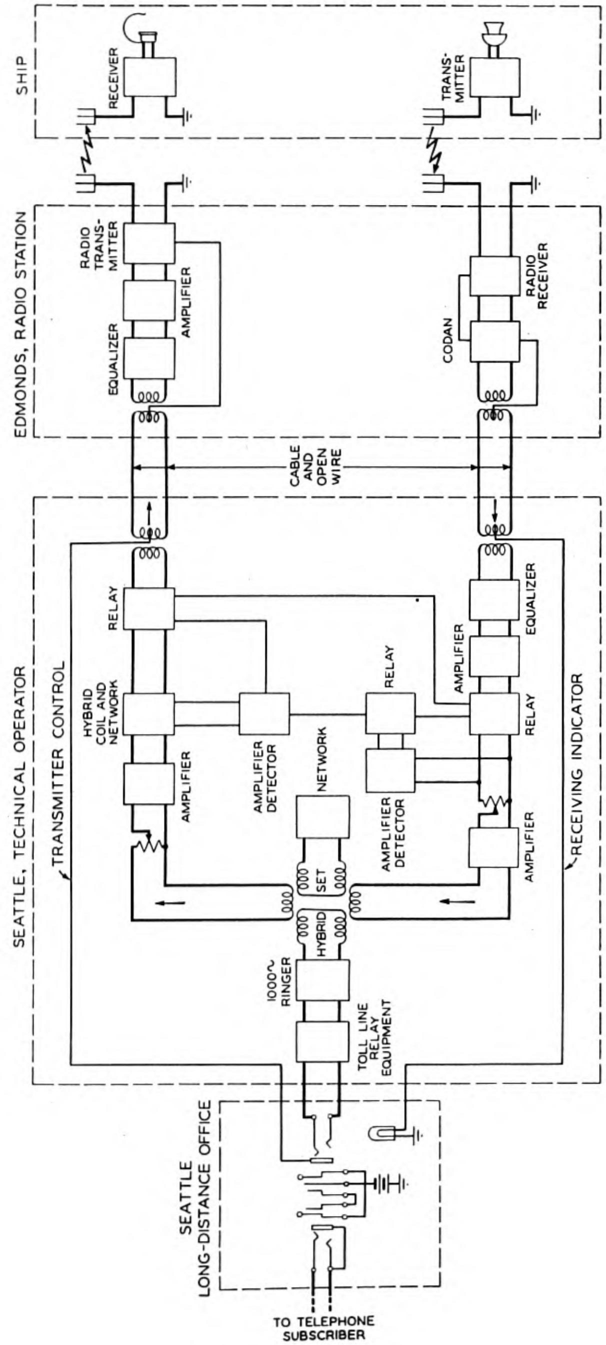


Fig. 2—Schematic of harbor radio telephone circuit.

circuit is monitored by means of an amplifier and loud speaker bridged across the circuit through a high impedance coil. This, of course, indicates to the attendant the general status of each connection as to transmission in both directions. Monitoring by means of a headset is also available and is usually resorted to in those instances where conditions require close attention to the adjustment of the apparatus to insure a satisfactory connection.

When a call originates on land, the toll operator connects the land line from the shore subscriber to the radio circuit, designated "Harbor," by means of a regular toll cord circuit. This operation, which requires the insertion of a plug in the jack of the radio circuit, automatically starts the transmitter at Point Edwards. It also indicates to the radio attendant by means of an alarm that a connection is being made and that his attention is required to ascertain whether any adjustments under his control are necessary for proper operation of the circuit. The switchboard operator then rings with a 1000-cycle signal which registers as an attention call to all ships which have their receivers operating. Selective ringing is available but requires that ships as well as station be properly equipped. The particular vessel to which a connection is to be made is called by name and station letter. This ship then starts its transmitter and reports and the two-way connection is established.

On calls originating from a ship, the boat's transmitter is energized and its carrier operates the codan circuit associated with the land station receiver. An auxiliary circuit actuated by the codan functions to signal the Seattle long distance operator by means of the regular toll line lamp signal. It also signals the radio attendant to stand by in the manner described above. The switchboard operator in responding to the signal inserts the answering plug of a cord circuit in the "Harbor" jack which energizes the land transmitter. Two-way telephone contact is then established with the ship, and from this point on the regular traffic operating procedure is followed in connecting the ship with the desired telephone station.

Ship Sets for Harbor Ship-to-Shore Service *

By H. N. WILLETS

A STUDY of conditions on board fishing trawlers and harbor craft indicates that the radio units must be small in size, rugged in construction and adaptable to remote control. Further, the units must permit operation by non-technical personnel and require a minimum of power for their proper functioning.

The striking similarity between these conditions and those already encountered in the furnishing of radio telephone equipment to aircraft permitted the use with small modification of a radio system already proved by years of use on this country's major airlines.

The ship transmitter, coded the 13-A, has an output of 50 watts. Its size is $13\frac{3}{4}'' \times 18\frac{7}{8}'' \times 10''$, its weight about 34 pounds.

The tube complement comprises a 5-watt audio-frequency amplifier tube, a 5-watt oscillator tube which is controlled by a quartz crystal oscillating at one-half the desired frequency and connected to the grid circuit of the oscillator tube, and a 50-watt screen grid tube as a first radio-frequency amplifier. The coupling from the oscillator is supplied by a radio-frequency transformer which freely passes the first harmonic of the quartz plate to drive the first radio-frequency amplifier. The output stage consists of two similar tubes in parallel to form a second radio-frequency amplifier. The two radio-frequency amplifier stages are coupled by a transformer which also acts as a band-pass filter and freely passes the output frequency. The two radio-frequency transformers are mounted in a single plug-in unit. The audio amplifier modulates the screen bias on the first and second radio-frequency amplifiers, giving substantially one hundred per cent modulation. The filament of each of the three radio-frequency amplifier tubes is in series with a ballast tube to protect them from fluctuations in the power supply.

The quartz crystal oscillator has the recently developed low temperature coefficient cut and the crystal holder is arranged with a heater which operates only should the temperature drop below zero Centigrade. The transmitter is arranged for one, two or three carrier frequencies, each requiring a separate crystal. If more than one frequency is provided, change from one to another is accomplished by a

* Digest of discussion prepared for Pacific Coast Convention, A.I.E.E., August, 1935.

single mechanical control. Remote selection of frequency may be by a flexible shaft or by a pull-wire arrangement.

The companion receiver, coded the 12-C, is a compact unit weighing about 16 pounds. It has a superheterodyne circuit employing six tubes and permits a quick shift between two fixed frequencies. High sensitivity and unusual selectivity are obtained. The antenna circuit is series tuned. The beating oscillator and modulator are combined in one tube while a quartz crystal assures the correct beating frequency. It is a plug-in unit similar to that used in the transmitter and ground to the carrier frequency plus the intermediate frequency of 385 kilocycles. Two stages of intermediate radio frequency are used. A duo diode triode tube is used as detector and first stage of audio-frequency amplification. The second audio and output tube is a pentode. Automatic volume control provides uniformity of output signal. A ballast tube protects the filaments against voltage fluctuations in the power supply.

Space is provided for two crystals and remote selection of either frequency may be provided by flexible shaft or pull wire. The controls are tied together through the receiver and transmitter mountings so that the frequency of both may be shifted simultaneously from a remote point.

While in the aircraft system the air pilot has his receiver always in use as he checks periodically on his route, it is obvious that the marine pilot would not require this constant contact with the shore station. To eliminate the necessity of loud speaker monitoring, selective ringing is available. It works from the output of the radio receiver with a conventional type selector which is stepped up in response to the impulses received from a dial actuating the shore transmitter. With this arrangement, any particular boat may be called, a bell announcing the incoming call.

For convenience of installation, the transmitter, receiver and selective ringing unit are mounted in a small metal cabinet. The cabinet provides protection to these units and incorporates the remote frequency change equipment and a junction box. All the connections between the unit mountings in the cabinet are permanent and terminate in the junction box. Provision is made in the base of the cabinet for two cables, one to the control equipment and the other to the power supply. The cabinet is provided with rubber feet which permit its installation in places subject to considerable vibration. One of the features of the installation is the ready access to any of the units for maintenance and test. The top and front of the cabinet may be removed and each unit, being "plug-in" mounted, is easily removable without disturbing connections. Adequate ventilation is provided as well as protection against direct splashing from above.

The power supply includes 12 volts for the filaments of the transmitter, receiver and selective signaling unit; 200 volts for the plate supply of the radio receiver, and 1,050 volts for plates of the tubes in the radio transmitter. For efficient operation these voltages must be held to reasonably close limits and experience has shown that wide

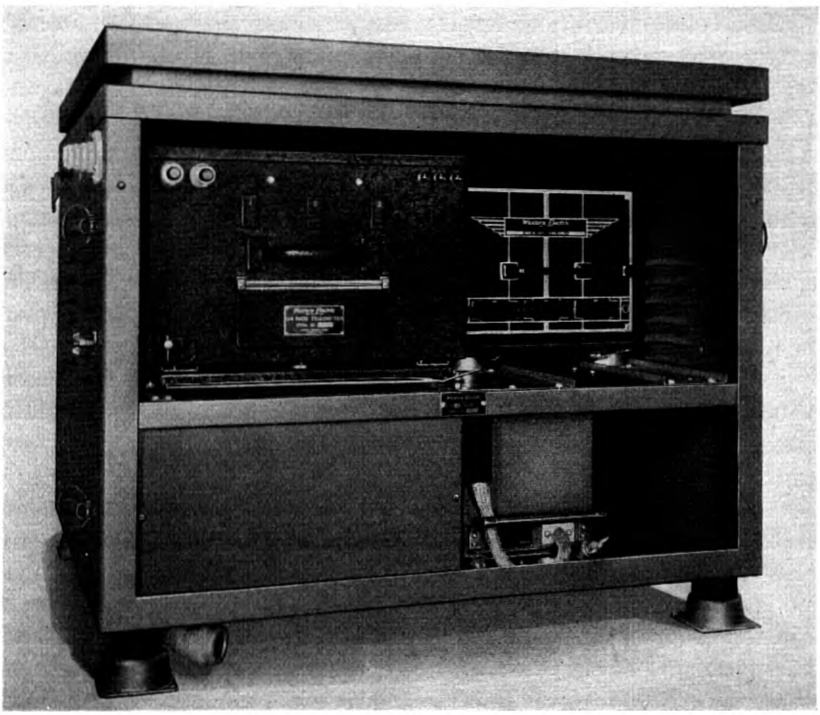


Fig. 1—Cabinet housing radio equipment; front panel removed showing left top radio transmitter, right top radio receiver, ringing unit beneath.

variations in the boat's power supply are to be expected. Therefore, two forms of power supply are generally recommended:

1. Motor generators for both transmitter and receiver;
2. A 12-volt storage battery and charging unit with dynamotors for the plate supply of the transmitter and receiver.

In the case of motor generators, two machines are provided, each consisting of a driving motor and a double winding generator mounted on a common cast iron base. The machines are of splash proof construction. The driving motors depend upon the source of power of each installation. Where the variation of the d-c. supply is greater than ± 3 per cent, a speed regulator is provided.

The generators of the double-voltage type have both a high and a low-voltage winding. One supplies 10 amperes at 13 volts and 0.1 ampere at 200 volts to the receiver; the other 15 amperes at 13 volts and 0.35 ampere at 1,050 volts to the transmitter. The units are equipped with filters to prevent disturbance in the radio equipment.

When voltage variations on shipboard are extreme, a 12-volt storage battery may provide the filament source as well as operate both a 200-volt and a 1,050-volt dynamotor. The battery charger may be of the automatic type or arranged for periodic charging.

The control unit is designed for installation at any convenient location. It provides facilities for remote starting and stopping the radio apparatus. It consists essentially of a single master-control switch, a telephone handset, call bell, a volume control for the receiver in the handset, and an antennæ meter for visual indication of the transmitter operation. A small lamp on the unit indicates when the receiver is in operation. To call or answer a call, one operates the single master-control switch and removes the handset from its hook switch. This controls the power to the radio transmitter. A push button in the handset handle automatically switches from receive to transmit position.

As an adjunct to the two-way radio telephone system, a radio compass has been developed which is essentially composed of a highly sensitive radio receiver and of a loop and "sense-indicating" antenna, and is extremely simple to operate. The accuracy of the arrangement is in the order of one degree when 200 to 300 miles from the marine radio beacon. Taking bearings on marine radio beacons is facilitated by a scale on the receiver calibrated directly in kilocycles and a chart on the receiver panel listing the radio beacons. After tuning, the loop is rotated by a hand wheel until the needle of a center-reading meter points to zero. A unique arrangement of the antenna permits a scale on the shaft of the loop to indicate the azimuth between the ship's heading and the direction of the received radio beacon when the meter reads zero. As the loop is turned to the right or left, the needle on the meter likewise swings to the right or left. There is no doubt as to the "sense" of the bearing as the meter gives positive visual indication when the loop is rotated. By plotting the direction of the boat from two or more radio beacons on the nautical chart, the intersection of these plotted lines determines the ship's position.

The direction-finding receiver employs a superheterodyne circuit, covering the frequency band from 242 to 515 kilocycles. A single control tunes the entire frequency range providing coverage of all the marine beacon stations and all the ship telegraph frequencies. The power supply for the receiver is the same as that for the other radio units. A loud speaker may be provided for identifying the station and to facilitate tuning. It may also be used for taking a bearing under severe static conditions. The loop assembly consists of a statically shielded loop winding, a supporting pedestal with shaft, a mounting flange, a hand-wheel, an azimuth scale, a compensator, an indicating meter, slip rings, lamp and terminal strip.

An Electromechanical Representation of a Piezoelectric Crystal Used as a Transducer *

By W. P. MASON

THE equivalent electrical representation of a piezoelectric crystal when used as an element in an electrical circuit has been discussed by several investigators,[1] who have arrived at the circuit shown on Fig. 1. Apparently, however, no circuit has been evolved for repre-

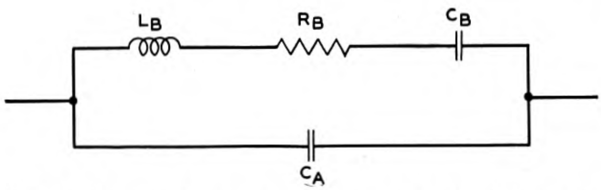


Fig. 1—Electrical representation of a piezoelectric crystal.

senting a crystal when it is used as a transducer to couple electrical circuits to mechanical systems. Since such crystals [2] are used in loud speakers, microphones, supersonic radiators, and other apparatus, it is a matter of importance to obtain such a representation. This paper discusses such an equivalent circuit and relates the elements to the mechanical, electrical, and piezoelectric constants of the material. When used as a purely electrical circuit, this representation reduces to that of Fig. 1.

When piezoelectric crystals are used to drive external mechanical systems, the modes of motion most often used are longitudinal vibrations perpendicular or parallel to the applied electric field. Accordingly, the elements of the equivalent network are derived for these cases only. The network can, however, represent any type of motion driving a load just as the network of Fig. 1 can represent the crystal for any type of motion.

Let us consider first the case of a crystal vibrating perpendicularly to

* Digest of an article appearing in the *Proc. I. R. E.* for October, 1935.

[1] W. G. Cady, *Phys. Rev.*, **XXIX**, 1 (1922); *Proc. I. R. E.*, **X**, 83 (1922). K. S. VanDyke, Ab. 52, *Phys. Rev.*, June, 1925; *Proc. I. R. E.*, June, 1928. D. W. Dye, *Proc. Phys. Soc.* (*London*), **XXXVIII**, (5), pp. 399–453. P. Vigoreaux, *Phil. Mag.*, December, 1928, pp. 1140–53.

[2] A. M. Nicolson, *Proc. A. I. E. E.*, **38**, 1315–1333, 1919. E. B. Sawyer, *Proc. I. R. E.*, **19**, No. 11, p. 2020, November, 1931. S. Ballantine, *Proc. I. R. E.*, **21**, No. 10, p. 1399, October, 1933.

the direction of the applied field. Two subdivisions of this case are usually of interest, the first when the crystal is supported at its center and drives two symmetrical loads, and the second when the crystal is supported on one end and drives a load on the other end. The symmetrical case is considered first.

By employing the well known analogies between electrical and mechanical systems, it is possible to obtain a simple network, expressed in terms of electrical symbols, which represents the properties of a piezoelectric crystal. In this representation, force is the analogue of voltage, mechanical displacement of electrical charge, and velocity of electrical current. When the electrodes are attached to the crystal faces, the equivalent network of the crystal is shown by Fig. 2, The voltage E

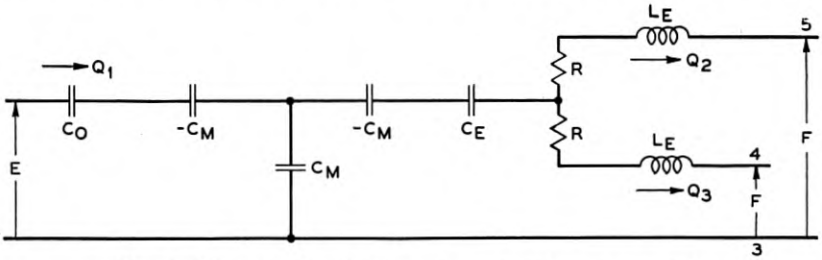


Fig. 2—Electromechanical representation of a symmetrical piezoelectric crystal.

is the voltage applied across the plates of the crystal, the force F is the force applied to each end of the symmetrical crystal, Q_1 is the electrical charge flowing in the wires connected to the crystal and Q_2 and Q_3 are the mechanical displacements of the ends of the crystal which are equal on account of the symmetry of the crystal.

The constants of the crystal can be evaluated by considering limiting cases. The capacitance C_0 is the electrostatic capacitance of the crystal clamped. The compliance C_E is the mechanical compliance of the crystal. In c.g.s. electrostatic and mechanical units, these have the values

$$C_0 = \frac{Kl_m l_0}{4\pi l_e}; \qquad C_E = \frac{sl_m}{l_e l_0}, \tag{1}$$

where K is the dielectric constant of the crystal clamped, l_e the dimension of the crystal in centimeters perpendicular to the surfaces of the electrodes, l_m the length of the crystal in the direction of vibration, l_0 the length of the third axis, and s the modulus of compliance of the crystal (the inverse of Young's modulus) along the axis of vibration. The inductance L_E represents the mass reaction of half the crystal. At low frequencies this will be equal to half the mass of the crystal, but at higher frequencies will be less due to the fact that the crystal does not move as a whole.[3] In order to resonate with the compliance C_E at the mechanical resonance frequency of the crystal, L_E must equal

$$L_E = \frac{2l_e l_m l_0 \rho}{\pi^2}, \tag{2}$$

where ρ is the density of the crystal. The resistances R shown include the dissipation due to internal friction, supersonic radiation from the ends of the crystal, friction at the point of support and all other sources of dissipation.

If F_0 is the force required to keep the crystal from expanding when an electric charge Q_1 is applied to the crystal then C_M the mutual capacitance-compliance of the crystal is equal to

$$C_M = Q_1/F_0. \tag{3}$$

Similarly if E_0 is the open circuit voltage for a given expansion $(Q_2 + Q_3)$ of the crystal then

$$C_M = \frac{Q_2 + Q_3}{E_0}. \tag{4}$$

In order to evaluate C_M in terms of the piezo-electric coefficient d, it is necessary to find the displacement for a free crystal when an electrical potential is applied to the crystal. Short circuiting the network of Fig. 2 on the mechanical ends and setting $F = 0$, we find

$$Q_2 + Q_3 = \frac{E\left[\dfrac{C_0 C_E}{C_M}\right]}{1 - \dfrac{C_0 C_E}{C_M^2}} = \frac{Ek\sqrt{C_0 C_E}}{1 - k^2}, \tag{5}$$

where k is the coefficient of coupling between the electrical and mechanical system is defined by the equation

$$k = \sqrt{C_0 C_E}/C_M. \tag{6}$$

Use is now made of the piezo-electric equation

$$e = dV, \tag{7}$$

[3] Strictly speaking the value of C_E is also a function of frequency, but at the first resonance it can be shown that it differs from its static value by the factor $8/[\pi^2 - k^2(\pi^2 - 8)]$ and $L_E = l_e l_m l_0 \rho/4$. For a highly coupled crystal, this factor does not differ much from unity, and hence in the interest of simplicity, the variations of C_E have been neglected.

where e is the strain (elongation per unit length) produced in the crystal by an applied potential gradient V. Comparing (5) with (7) and noting that $Q_2 + Q_3 = el_m$, and $V = E/l_e$, we have on the insertion of the values for C_0 and C_E from (1)

$$d = \frac{k\sqrt{\frac{Ks}{4\pi}}}{1 - k^2}. \tag{8}$$

Solving for k we find

$$k = \frac{1}{2d}\sqrt{\frac{Ks}{4\pi}}\left[-1 + \sqrt{\frac{1 + 16\pi d^2}{Ks}}\right] \doteq d\sqrt{\frac{4\pi}{Ks}}, \tag{9}$$

when $16\pi d^2/Ks$ is a small quantity as it is for quartz.

When the crystal is used as an element in an electrical network, and allowed to vibrate freely, the force F of Fig. 2 can be set equal to zero and the network short-circuited. Solving for the impedance on the electrical side we find

$$Z_C = \frac{-j(1 - k^2)}{2\pi f C_0}\left[\frac{1 - f^2/f_1^2 + j/q(1 - k^2)}{1 - f^2/f_2^2 + j/q}\right], \tag{10}$$

where $f_2^2 = f_A^2$; $f_1^2 = f_A^2(1 - k^2)$ (f_A being the natural mechanical resonance frequency of the crystal), and q is the ratio of the reactance of the condenser C_E to the resistance $R/2$ or

$$q = 2/2R\pi f_A C_E. \tag{11}$$

It is easily shown that the impedance Z_c is also the impedance of the network of Fig. 1 if [4]

$$\begin{aligned} C_A &= C_0; \\ C_B &= C_0 k^2/(1 - k^2); \\ L_B &= 1/4\pi^2 f_A{}^2 k^2 C_0; \\ R_B &= 1/2\pi f_A C_0 k^2 q. \end{aligned} \tag{12}$$

Hence the representation in Fig. 2 reduces to the well known Fig. 1, when the crystal is free to vibrate.

A network representing the second case, when one end is supported with the other end used to drive a load, is shown on Fig. 3. The method of deriving the constants is the same and all of the constants

[4] If account is taken of the variation of C_E and L_E with frequency, the elements are $C_A = C_o$; $C_B = \dfrac{8k^2C_o}{\pi^2(1 - k^2)}$; $L_B = \dfrac{1}{32k^2 f_A{}^2 C_o}$; $R_B = 1/2\pi f_A C_o k^2 q$ where $q = \dfrac{1}{\pi R C_E f_A}$.

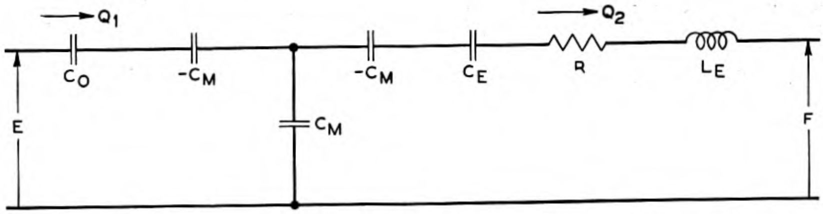


Fig. 3—Electromechanical representation of a piezoelectric crystal clamped on one end.

are the same except L_E, which is twice as large since twice the mass is moved from the clamping position.

When the direction of motion is parallel to the direction of the applied field, the same networks hold but the elements have different lengths entering into their determinations. The direction of the applied field and of the motion is designated by l_e. The other two axes are still designated by l_m and l_0. The resulting constants are

$$C_0 = \frac{Kl_m l_0}{4\pi l_e}; \quad C_E = \frac{sl_e}{l_m l_0}; \quad C_M \doteq \frac{Ks}{4\pi d}; \quad L_{E_1} = \frac{2\rho l_e l_0 l_m}{\pi^2}; \quad L_{E_2} = \frac{4\rho l_e l_0 l_m}{\pi^2}, \tag{13}$$

where L_{E_1} is the mechanical inductance for the symmetrical case (Fig. 2) and L_{E_2} for the dissymmetrical case (Fig. 3).

A simple example of the use of Fig. 2 in determining the effect of a mechanical load on the impedance of a crystal is the problem of finding out how much the energy radiated by a crystal to the surrounding air affects the decrement of a quartz crystal vibrating longitudinally. When a crystal vibrates, energy is radiated to the surrounding medium by the motion of the ends of the crystal. If the dimensions of the ends of the crystal are comparable to a wave-length or greater—which they will ordinarily be for a crystal vibrating at a high frequency—it is well known [5] that the radiating surface experiences a resistance to motion equal to

$$R_R = \rho_A b \text{ (mechanical ohms)} \tag{14}$$

per square centimeter, where ρ_A is the density of the medium and b the velocity of propagation. For air R_R is about 41 ohms per square centimeter. Hence the equivalent circuit for a crystal vibrating in air is Fig. 2 terminated at the terminals 3–4 and 3–5 by the mechanical

[5] See "Theory of Vibrating Systems and Sound," I. B. Crandall, Chap. 4, D. Van Nostrand Co.

resistances

$$R_R = 41 l_e l_0. \tag{15}$$

If all other sources of dissipation were eliminated, the radiation resistance would produce a limiting value for the decrement of a crystal which may be calculated as follows. From equation (11), the value of q for the mechanical system is

$$q = \frac{1}{\pi f_A R_R C_E} = \frac{2\sqrt{\rho/s}}{41\pi} \tag{16}$$

on inserting the values of R_R, C_E, and $f_A = 1/2l_m\sqrt{\rho s}$. Since $\rho = 2.65$ and $s = 1.27 \times 10^{-12}$ for a perpendicularly cut quartz crystal

$$q = 2.24 \times 10^4. \tag{17}$$

The decrement of a crystal in terms of the circuit of Fig. 1 is

$$\delta = \frac{R_B}{2 f_A L_B}. \tag{18}$$

Inserting the values of footnote (4) we find

$$\delta = \frac{8}{\pi q} = 1.14 \times 10^{-4}. \tag{19}$$

Van Dyke [6] has measured the limiting value of the decrement of a perpendicularly cut quartz crystal vibrating in air and finds it to be 1.26×10^{-4}. Since the residual losses were about 5 per cent of the radiation losses, this agrees well with the value found in equation (19).

The equivalent circuits of Figs. 2 and 3 may also be used as the basis of design for mechanical systems, such as loud speakers, microphones and supersonic radiators.

[6] *Proc. I. R. E.*, April, 1935.

Abstracts of Technical Articles from Bell System Sources

Some Methods for Making Resonant Circuit Response and Impedance Calculations.[1] H. T. Budenbom. This article presents a series of short-cut methods for the computation of the amplitude and phase-response characteristics of one- and two-mesh circuits; the extension of the method to three- and four-mesh networks is carried out formally. It also treats the impedance-frequency characteristics of singly resonant circuits. The plan of attack in the single-circuit impedance and multi-mesh response cases is to express the desired circuit or transfer impedance as a polar numeric multiplied by a sizing constant which turns out to be a simple reactance element of the circuit.

A Rapid Method for the Determination of Sulfur in Ferromagnetic Alloys.[2] B. L. Clarke, L. A. Wooten, and C. H. Pottenger. A method is described for the determination of sulfur in ferrous alloys of high-nickel content, wherein the sample is heated in hydrogen at 1100° to 1200° C., and the liberated hydrogen sulfide absorbed in ammoniacal cadmium nitrate solution. The method is generally applicable to iron-nickel-cobalt alloys.

The precision of the method for the range of 0.005 to 0.020 per cent sulfur is shown to be $\pm$ 0.001 per cent sulfur on the basis of a 10-gram sample.

The Newly Discovered Elementary Particles.[3] Karl K. Darrow. During the three years following the discovery of heavy hydrogen in the latter part of 1931, four additional elementary particles of matter lighter than the alpha particle (nucleus of the ordinary helium atom) have been discovered where only two were known previously. This is considered extraordinary even for the present rapid pace of physics which sometimes requires physicists to revise their fundamental concepts of matter almost overnight. These five newly discovered particles formed the subject matter of a highly interesting and instructive address delivered recently by Doctor Darrow at a meeting of the American Institute of Electrical Engineers' New York Section; the essential substance of this illuminating address is presented in the published article.

[1] *Radio Engg.*, August, 1935.
[2] *Indus. & Engg. Chem.*, Analytical Edition, July 15, 1935.
[3] *Elec. Engg.*, August, 1935.

Magnetic Hysteresis at Low Flux Densities.[4] W. B. ELLWOOD. The energy loss per cycle in a ferromagnetic material subjected to small alternating fields is sometimes separated into three parts: the first due to eddy current loss, the second to hysteresis presumed to follow Rayleigh's laws at these flux densities; the origin of the remainder is unknown and is the subject of much controversy; it has been variously termed "magnetic viscosity," "after effect," and "square law hysteresis." In studying energy loss in a ring of compressed iron dust, hysteresis loops have been measured ballistically by a new method with a relative error in B_m as low as 0.01 per cent. The range in maximum flux density is from 2 to 100 gauss. The loops are lenticular and very slender, B_m being about 1500 times the remanence for the smallest loop. The smaller loops are at flux densities considerably below those investigated by Rayleigh. His findings as to variation of area, of remanence, and of permeability with loop amplitude are confirmed and extended through the new range of flux densities, though the shape of the loops is not as simple as that he proposed. The energy loss per cycle is proportional to B_m^3 while the remanence is proportional to B_m^2. The ballistic measurements have been compared with a-c. bridge measurements on the same specimen. The loss of unknown origin is not included in the hysteresis loss measured ballistically. Comparison is made between the third harmonic induced voltage computed by a Fourier series analysis of the ballistic loops and the harmonic actually generated by the specimen. Agreement is found between the measured and the computed values. Possible explanations of the discrepancy between the ballistic observed energy losses and a-c. findings are discussed.

A Fugue in Cycles and Bels.[5] JOHN MILLS. Engineers who are questioned by their musical acquaintances about electrical transmission and what it is likely to do to the art are likely to find an explanation, even in the simplest of terms, shooting over the heads of their audience. The reason lies not in the inherent difficulty of the concepts, but in their number, which exceeds the power of memory to retain as unrelated facts. In this volume John Mills has strung his facts together on the thread of logical relationship, but he has tied them into his readers' existing knowledge by many a deft touch of anecdote or humor. Frequency, with its relation to harmony and discord, opens the volume; there follows a physical picture of vibrations in various media and how one form is transformed into another. Pitch and intensity are next considered. An entire section is devoted to tele-

[4] *Physics*, July, 1935.
[5] Published by D. Van Nostrand Company, Inc., New York, N. Y., 1935.

phonic studies of hearing; it includes such topics as the ear, the amplifier, transmission, loudness, overloading and distortion.

A third section, which discusses the electrical future for music, will most intrigue the musician. After surveying the present state of the art of pick-up, transmission, recording and reproduction of music, the author visualizes some of the possibilities which might be most immediately realized: reproduction in auditory perspective, electrical music and electrical aids in teaching.

To permit a rapid survey by the general reader, Mr. Mills has withdrawn all tables and graphs from the main text and has grouped them, with necessary explanations, at the end. This section forms a useful compendium of numerical information on frequencies, decibels, thresholds, response curves of microphones and loud speakers, loudness and energy levels, masking and the like.

The engineer will find "A Fugue in Cycles and Bels" interesting and easy to read, and a source of data not always available, and he can recommend it to those musicians whose serious interest in their art leads them to delve into its physical basis.

Influence of Experimental Technique on the Measurement of Differential Intensity Sensitivity of the Ear.[6] H. C. MONTGOMERY. The lack of agreement among previous measurements of differential intensity sensitivity indicates that the values obtained depend to a large extent on the experimental conditions. The relative importance of various factors is indicated, and a procedure is suggested which was designed to give the smallest possible values of differential intensity sensitivity. Intensive measurements made by this method upon a single subject, using a pure tone of 1000 cycles, gave values consistently smaller than any previously reported. There is no sharp division between intensity changes which can be perceived and those which cannot. The response of the subject is essentially variable and can be described only by statistical methods.

Diurnal and Seasonal Variations in the Ionosphere During the Years 1933 and 1934.[7] J. P. SCHAFER and W. M. GOODALL. The most important results of daily ionospheric measurements made at Deal, New Jersey, latitude 40° 15′ N., longitude 74° 02′ W., over the period from March, 1933, to May, 1934, are given in this paper and may be summarized as follows:

1. There was a definite correlation between the noon value of ionic density of the F_1 region and magnetic disturbances, a decrease in ionic density being obtained on magnetically disturbed days.

[6] *Jour. Acous. Soc. Amer.*, July, 1935.
[7] *Proc. I. R. E.*, June, 1935.

2. The noon value of ionic density of the E and F_1 region attained a maximum in summer and a minimum in winter whereas the reverse condition, of minimum in summer and maximum in winter, was found for the F_2 region.

3. The time of maximum ionic density of the F_2 region varied with the seasons of the year, occurring near noon in winter and near sunset in summer.

The paper also shows a series of virtual height contour maps for the four seasons of the year.

Utilization of Electrical Resistance Measurements for Controlling the Composition of Alloys.[8] E. E. SCHUMACHER and L. FERGUSON. To stimulate interest in the more general use of the electrical resistance of solid solutions for indicating the composition of alloys, details concerning one application are set forth in this paper. While for other systems certain elements of the method as outlined here might have to be modified, it appears that the general principle could be advantageously utilized in many instances. This paper presents, therefore, a practical resistance method of supplementing ordinary chemical analyses in the determination of the antimony content of lead-antimony alloys ranging at least from 0.4 to 1.1 per cent. It has been found that, within this range, a 1 per cent change in resistance is produced by a change in antimony content of less than 0.1 per cent and that the maximum deviation in resistance data may reasonably be made so small as to correspond to not more than $\pm$ 0.01 per cent antimony.

[8] *Metals and Alloys*, June, 1935.

Contributors to this Issue

Leonard Gladstone Abraham, B.S., 1922, M.S., 1923, University of Illinois. American Telephone and Telegraph Company, Department of Development and Research, 1923–34; Bell Telephone Laboratories, 1934–. Mr. Abraham has been engaged in transmission development work on toll telephone systems.

Austin Bailey, A.B., University of Kansas, 1915; Ph.D., Cornell University, 1920; Instructor in Physics, Cornell University, 1915–18; Signal Corps, U. S. A., 1918–19; Assistant Professor of Physics, University of Kansas, 1921–22. American Telephone and Telegraph Company, Department of Development and Research, 1922–34; Bell Telephone Laboratories, 1934–. Dr. Bailey's work has been largely along the line of methods for making radio transmission measurements and of long-wave radio problems.

F. A. Gifford, B.S., Tufts College, 1920. General Electric Company, Lynn, Massachusetts, 1920–22; American Radio and Research Corporation, Medford, Massachusetts, 1922; New England Telephone and Telegraph Company, Engineering Department, Boston, 1922–. Mr. Gifford has been engaged on various types of telephone engineering work, devoting a large part of his time during the past four years to problems in connection with the Boston Marine Radio Telephone Service.

Earl B. Hansen, B.S. in Electrical Engineering, University of California, 1920. Pacific Telephone and Telegraph Company, 1920 to present date. Latterly, Mr. Hansen's duties have been in the field of toll transmission.

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. R. Huntley, B.S., University of Wisconsin, 1921. Wisconsin Telephone Company, Engineering Department, 1917–30; American

Telephone and Telegraph Company, Department of Operation and Engineering, 1930–. Mr. Huntley's work has been concerned principally with transmission and inductive coordination matters.

Dr. Jewett, as Vice President of the American Telephone and Telegraph Company and President of the Bell Telephone Laboratories, needs no introduction to *Technical Journal* readers.

S. A. Levin, E.E., Chalmers Technical Institute, Gothenburg, 1919; Technische Hochschule, Berlin, 1920–21; Technische Hochschule, Dresden, 1921–23. Radio Department, General Electric Company, Schenectady, N. Y., 1923–26; Engineering Department, National Electric Light Association, New York, N. Y., 1926–30; Bell Telephone Laboratories, 1930–. Mr. Levin's work has to do with the development of high-frequency measuring equipment for carrier systems.

Frederick B. Llewellyn, M.E., Stevens Institute of Technology, 1922; Ph.D., Columbia University, 1928. Western Electric Company, 1923–25; Bell Telephone Laboratories, 1925–. Dr. Llewellyn has been engaged in the investigation of special problems connected with radio and vacuum tubes.

W. P. Mason, B.S. in Electrical Engineering, University of Kansas, 1921; M.A., Columbia University, 1924; Ph.D., 1928. Bell Telephone Laboratories, 1921–. Dr. Mason has been engaged in investigations on carrier transmission systems and more recently in work on wave transmission networks, both electrical and mechanical.

R. B. Meader, B.S., University of New Hampshire, 1921. New England Telephone and Telegraph Company, 1922–. Since 1930 Mr. Meader has worked chiefly on radio telephone field strength surveys and other activities leading to the establishment of the Green Harbor Radio Telephone Station and the development and operation of the Boston Marine Radio Telephone Service.

E. J. O'Connell, B.S., Northwestern University, 1924. American Telephone and Telegraph Company, Long Lines Department, 1924–25; Illinois Bell Telephone Company, Engineering Department, 1925–28; American Telephone and Telegraph Company, Department of Operation and Engineering, 1928–. Mr. O'Connell's work has been concerned principally with inductive coordination matters.

Liss C. Peterson, E.E., Chalmers Technical Institute, Gothenburg, 1920; Technische Hochschule, Charlottenburg, 1920–21; Technische Hochschule, Dresden, 1921–22; Signal Corps, Swedish Army, 1922–23.

American Telephone and Telegraph Company, 1925–30; Bell Telephone Laboratories, 1930–. Mr. Peterson is engaged in the study of modulation and other problems connected with high-frequency carrier systems.

G. S. Phipps, B.S. in Electrochemical Engineering, Pennsylvania State College, 1930. Bell Telephone Laboratories, 1930–. Mr. Phipps has been engaged principally in the metallurgical investigation of lead, aluminum and zinc alloys.

E. E. Schumacher, B.S., University of Michigan; Research Assistant in Chemistry, 1916–18. Engineering Department, Western Electric Company, 1918–25; Bell Telephone Laboratories, 1925–. As Assistant Research Metallurgist, Mr. Schumacher is in charge of a group whose work relates largely to research studies on metals and alloys.

U. B. Thomas, Jr., B.S. in Chemistry, William and Mary, 1929. Bell Telephone Laboratories, 1929–. Mr. Thomas has been engaged principally in the study of base metal contacts, storage batteries, and allied electrochemical problems.

Howard M. Thomson, B.S. in E.E., University of Washington, 1930. Bell Telephone Laboratories, 1930–32; American Telephone and Telegraph Company, Department of Development and Research, 1932–34; Bell Telephone Laboratories, 1934–. Mr. Thomson's work has been principally on studies of long-wave radio transmission and field tests on an experimental transmitting wave antenna.

H. N. Willets, Western Electric Company, Engineering Inspection Department, 1917–25; Philadelphia Instrument Shop, 1921–25; Supply Department, 1925–26. Graybar Electric Company, 1926–28; Western Electric Company, 1928–. Mr. Willets is engaged in the promotion of aircraft radio, broadcasting, and ship-to-shore radiophone and harbor craft radiophone.